(2)

Washington
WASHINGTON UNIVERSITY IN ST. LOUIS

FINAL REPORT

HIGH-RESOLUTION RADAR IMAGING

O.N.R. Contract N00014-86-K-0370

15 January 1989 - 14 January 1990

Principal Investigator: Donald L. Snyder
Address: Electronic Systems and Signals Research Laboratory
Campus Box 1127
Washington University
One Brookings Drive
St. Louis, MO 63130
Telephone: (314) 889-6159

Scientific Program Director: Dr. Rabinder Madan
Address: Office of Naval Research
Code 1114SE
800 North Quincy Street
Arlington, Virginia 22217-5000

DISTRIBUTION

	copies
Mr. John W. Michalski Office of Naval Research Resident Representative Federal Building, Room 286 536 South Clark Street Chicago, Illinois 60605-1588	1
Dr. Rabinder Madan Office of Naval Research Code 1114SE 800 North Quincy Street Arlington, Virginia 22217-5000	1
Director Naval Research Laboratory Attn.: Code 2627 Washington, DC 20375	1
Defense Technical Information Center Building 5 Cameron Station Alexandria, Virginia 22314	12
Mr. Harper J. Whitehouse Naval Ocean Systems Center Code 7402 San Diego, California 92152	1

STATEMENT "A" per Rabinder Madan
ONR/Code 1114SE
TELECON 4/12/90 VG

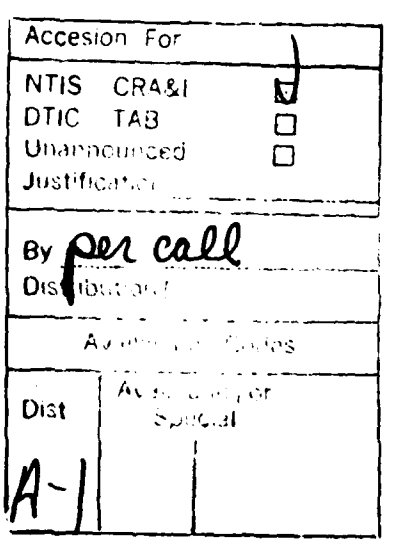


Table of Contents

1. Introduction 1

2. Summary of Work Accomplished 1
2.1 Conventional Approach to Imaging 1
2.2 Estimation-Theory Approach to Imaging 2

3. Other Project Activities 6
3.1 Optical Radar Workshop 6
3.2 Invited Presentations 7

4. Work in Progress 7
4.1 Real-Data Experiment 7
4.2 Graduate-Student Theses 8

5. Personnel 8

6. Appendices 11

6.1 Reprint of: D. Snyder, J. O'Sullivan, and M. Miller, The Use of Maximum Likelihood Estimation for Forming Images of Diffuse Radar Targets from Delay-Doppler Data, *IEEE Trans. on Information Theory*, Vol. 35, pp. 536-548, 1989. 11

6.2 Reprint of: P. Moulin, D. Snyder, and J. O'Sullivan, "Maximum-Likelihood Spectrum Estimation of Periodic Processes from Noisy Data," Proc. 1989 CISS Conference, Johns Hopkins University, Baltimore, MD, March 1989. 25

6.3 Reprint of: J. A. O'Sullivan, D. L. Snyder, and P. Moulin: "The Role of Spectrum Estimation in Forming High-Resolution Radar Images", Proc. ICASSP 1989, Glasgow, U.K., May 1989. 37

6.4 Reprint of: P. Moulin, D. L. Snyder, and J. A. O'Sullivan: "A Sieve-Constrained Maximum-Likelihood Estimator for the Spectrum of a Gaussian Process", Proc. 28th Allerton Conference, Urbana-Champaign, IL, Sept. 1989. 41

6.5 Reprint of: J. A. O'Sullivan, P. Moulin, D. L. Snyder, and S. P. Jacobs, "Computational Considerations for Maximum-Likelihood Radar Imaging," Proc. 1990 CISS, Princeton University. 51

6.6 Preprint of: J. A. O'Sullivan and D. L. Snyder, "High Resolution Radar Imaging Using Spectrum Estimation Methods," Proc. August 1989 Program on Signal Processing, Institute for Mathematics and Its Applications, University of Minnesota, Minneapolis MN, to appear. 65

1. Introduction

This final report contains a summary of work accomplished on O. N. R. Contract N00014-86-K-0370, *High-Resolution Radar-Imaging*, for the period from 15 January 1989 to 14 January 1990. Also included is a description of research in progress that will be phased out as funding for this project has been terminated.

The goal of this project is to formulate and investigate new approaches for forming images of radar targets from spotlight-mode, delay-doppler measurements. These measurements can be acquired with a high-resolution radar-imaging system operating with an optical- or radio-frequency carrier. Work in this reporting period has concentrated on our *estimation-theory approach* to forming high resolution images. This approach accounts for measurement noise and for the statistical properties of radar-backscatter data.

2. Summary of Work Accomplished

2.1 Conventional Approach to Imaging

A computer implementation of the conventional method for forming radar images via the two-dimensional Fourier-transform has been implemented by Mr. D. Porter, who is an undergraduate student in Electrical Engineering. This is used to compare and evaluate images produced conventionally with those produced using the new methods we are studying. There are two modules in the program.

The first module produces simulated radar back-scatter data. The simulation calculates ideal samples of a received signal when a stepped-frequency waveform is reflected off of a target having a specified scattering function and the received signal is mixed with the transmitted signal and then sampled in quadrature to produce a complex-valued sample. The input of this simulation includes the scattering function of the target, the desired resolution, and the base frequency of the transmitted waveform. Complex samples of the received signal form the output of the program. The computational complexity is $O(N^4)$, where N is the number of resolution cells in one target dimension. While coding efficiencies have been exploited to a high degree, it may still be desirable to effect a parallel implementation of the calculations for producing results more quickly for high resolution images when N is large.

The processing of simulated and real-data samples is performed by the second module, which applies a slightly modified two-dimensional DFT to the samples. The squared magnitude of the resulting transform is then produced as the target's image. It is assumed that scatterers do not migrate out of their resolution cells over the time of data collection, which is an approximation. When this approximation is not well met, the resulting image is distorted due to range walk. A number of ways are described in the literature to deal with this problem. First of all, the transmitted waveform's base frequency can be chosen sufficiently high so that the desired cross range resolution can be achieved without requiring a wide variation in view angle over the data collection interval. Since only a narrow angle is required, scatterers will remain within the confines of their original resolution cell. Secondly, the effect of range walk can be seen as equivalent to collecting polar-formatted data in the spatial frequency domain, yet transforming it as if it were in a rectangular format. Interpolation using, for example, cubic splines in order to reposition the data samples onto a rectangular grid in the spatial frequency domain is a method of "focusing" the image, thereby reducing the error due to range walk. This focusing method, which we are using, is described by D. Wehner [*High Resolution Radar*, Artech House, pp. 311-317, 1987].

In addition to the two main modules of the program, additional utility modules have been developed for the generation of scattering function files, the display of images on MASSCOMP and SUN workstations, and the conversion of data between the different formats used on the different computers used in our study.

2.2 Estimation-Theory Approach to Imaging

Significant progress has been made on the method we are developing for producing images of low visibility targets modeled as a diffuse scattering object. The problem of estimating the scattering function of a diffuse target is ill posed, with the result that estimates are unstable having a the rough appearance of an object with strong speckle noise. Thus, regularization is required in order to stabilize estimates of the scattering function, which is a two-dimensional power-density spectrum. Of particular importance in our work over the reporting period is the identification of a method for regularizing estimates via the method of sieves introduced

by U. Grenander (*Abstract Inference*, Wiley, 1981). We expect that this will be an important development not just for radar imaging but also for any problem where a power-density spectrum must be estimated from noisy data. The asymptotic properties of this method of regularization have also been instrumental in the identification of a computational method for producing target images practically. The inversion of large matrices is not required under some conditions that are met is practice, which removes a major impediment previously existing with our method.

The following paragraphs contain the abstracts of publications describing our results. Details are given in the appendices, which contain reprints of the publications.

2.2.1 Abstract of: D. Snyder, J. O'Sullivan, and M. Miller, "The Use of Maximum Likelihood Estimation for Forming Images of Diffuse Radar Targets from Delay-Doppler Data," IEEE Trans. on Information Theory, Vol. 35, pp. 536-548, 1989; see Appendix 6.1 for a reprint.

This publication gives the model and fundamental estimation equations for the method we are developing. The abstract is:

> "A new approach to high resolution radar imaging is presented. The starting point is a model of the radar echo signal based on the physics governing radar reflections. This model has been used several times in the past for describing radar targets that are rough compared to the wavelength of the transmitted radiation. Without specifying precisely what the transmitted signal is, a general estimation-based procedure is derived for obtaining images. After discretizing the model, the radar imaging problem reduces to the task of estimating discretized second-order statistics of the reflectance process of the target. Maximum likelihood estimates of these statistics are obtained as the limit point of an expectation-maximization algorithm."

2.2.2 Abstract of: P. Moulin, D. Snyder, and J. O'Sullivan, "Maximum-Likelihood Spectrum Estimation of Periodic Processes from Noisy Data," Proc. 1989 CISS Conference, Johns Hopkins University, Baltimore, MD, March 1989; see Appendix 6.2 for a reprint.

"We have developed a new approach to maximum-likelihood spectrum estimation of wide-sense stationary processes from noisy data. A statistical model for the data is defined. The process whose spectrum is sought is wide-sense stationary, periodic and Gaussian, and its observations are corrupted by an additive white noise [and a linear transformation]. [For our radar-imaging problem, the Gaussian process models radar back-scatter data from a diffuse target, the spectrum is the target's scattering function, the data are corrupted by additive noise, and the linear transformation depends on the transmitted signal.] A maximum-likelihood formulation of this problem has been derived, and the equations are solved numerically via the expectation-maximization algorithm. This approach presents several attractive features, an important one being that the noise corrupting the observations is now taken into account.

We present some recent developments for this problem. The statistical performance of the new maximum-likelihood spectrum estimator is studied both theoretically and numerically. Comparison with traditional estimators, such as the periodogram, highlight several strong points of the method. We also identify certain limitations, namely the instability of estimates for high noise levels, [which is due to the ill-posed nature of the spectrum estimation problem]. These limitations can be alleviated if *a priori* information about the signal is available. Two such problems are discussed [in Appendix 6.2] in which the information at hand has the form of a constraint on the input signal-to-noise ratio.

We show [in Appendix 6.2] how such information can be incorporated in the maximum-likelihood estimation procedure. First we assume the signal power to be known. Theoretical issues of existence and uniqueness of the solution are discussed.

We proceed with a problem in which the information is less complete, when only an upper bound and/or a lower bound on the signal power are available. The statistical performance of both constrained estimators is quantitatively studied."

2.2.3 Abstract of: J. A. O'Sullivan, D. L. Snyder, and P. Moulin: "The Role of Spectrum Estimation in Forming High-Resolution Radar Images", Proc. ICASSP 1989, Glasgow, U.K., May 1989; see Appendix 6.3 for a reprint.

"We have developed a new approach to forming high-resolution images of radar targets from delay-doppler, spotlight-mode radar data. This approach is based on a model for the target's reflectivity in terms of wide-sense stationary, uncorrelated scatterers having complex-valued Gaussian statistics. The imaging problem is to estimate the target's scattering function in terms of radar-echo data acquired with a series of target illuminations. We develop [in Appendix 6.3] a method for solving this multidimensional spectrum estimation problem through the use of maximum-likelihood estimation implemented via the expectation-maximization algorithm."

2.2.4 Reprint of: P. Moulin, D. L. Snyder, and J. A. O'Sullivan: "A Sieve-Constrained Maximum-Likelihood Estimator for the Spectrum of a Gaussian Process", Proc. 28th Allerton Conference, Urbana-Champaign, IL, Sept. 1989; see Appendix 6.4 for a reprint.

"Maximum-likelihood spectrum estimation is an ill-posed problem. In this paper, we use of a method of sieves for addressing this issue. The estimate of the spectrum is constrained to a subset of some Hilbert space of functions over which a complete set of nonorthogonal basis functions is defined. The estimate is then represented by a countable set of coefficients in a nonorthogonal series expansion. By defining an appropriate sieve on this countable set, our problem reduces to maximum-likelihood estimation of the parameters in the sieve. Three main attractive features of this approach are: (1) the nonorthogonal expansion is a convenient framework for defining the sieve and including *a priori* information; (2) mean-square consistency of the estimates can be expected; and (3) we have derived a tractable alternating maximization algorithm for estimating the parameters. The setup of this

problem is general and can be applied without major difficulties to the estimation of higher-dimensional spectral functions, as occurs, for example, in imaging radar targets from delay-doppler data."

2.2.5 Abstract of: J. A. O'Sullivan, P. Moulin, D. L. Snyder, and S. P. Jacobs, "Computational Considerations for Maximum-Likelihood Radar Imaging," Proc. 1990 CISS, Princeton University; see Appendix 6.5 for a reprint.

"Recent papers have outlined a new approach for spectrum estimation and radar imaging based on expectation-maximization algorithms for structured covariance estimation. Performance of this approach has been promising for the problems studied. Application of this approach to real data sets has been limited, however, due to the need to invert a matrix whose dimension equals the size of the data set. For radar applications where an image is to be formed, data sets can be on the order of 2^{14} for 128×128 images. This makes the use of the new approach difficult in its previously described form. This paper proposes both approximation methods for inverting typical matrices and constraints on radar transmitted signals which make maximum likelihood image estimation viable. These constraints may be satisfied for real signals used in radar imaging systems. Simulations are shown to demonstrate the performance of the algorithms. Finally, motivated by the images resulting from the simulations, regularization methods are discussed."

3. Other Project Activities

3.1 Optical Radar Workshop

At the request of Dr. W. J. Miceli, Office of Naval Research, Boston MA, we organized and hosted a one-day workshop on laser radar imaging on April 12, 1989. The purpose of the meeting was to discuss various tomographic image reconstruction methods, their applicability to laser radar imaging, and their implementation via a suitable real-time signal processing system. The goal was to provide technical interaction among researchers interested in the topic.

Most participants were funded through O.N.R. and/or SDIO/T/IS. The approximately twenty attendees were from government laboratories and organizations, university research laboratories, and industry.

3.2 Invited Presentations

In August 1989, a special program on the subject of signal processing was held at the Institute for Mathematics and Its Applications, of the University of Minnesota in Minneapolis. We were invited to present our work on the radar imaging problem and to prepare a chapter for a book to be published by the I. M. A. This chapter, coauthored by J. O'Sullivan and D. Snyder and titled "High Resolution Radar Imaging Using Spectrum Estimation Methods," will appear during 1990; a preprint is in Appendix 6.6. As a result of the interest in radar detection and imaging problems that developed during this program, a second program on the topic of radar and sonar has been organized and will take place in June 1990 at the I. M. A.

4. Work in Progress

4.1 Real-Data Experiment

An effort to collect real data with which to test and compare our methods for radar imaging has been initiated in collaboration with the McDonnell-Douglas Company in St. Louis. A sphere having a diameter of one meter and a rough surface was placed on a rotating pedestal in a compact radar test-range, and data were collected at several values of signal-to-noise ratio. This object was selected because of its simplicity and the fact that its scattering function can be predicted analytically, thereby providing a test object having known characteristics with which to compare results. These data have only recently been acquired and have not yet been processed.

4.2 Graduate-Student Theses

4.2.1 Pierre Moulin

Pierre Moulin is presently writing his doctoral dissertation on the subject of estimation methods for forming images of diffuse radar targets. It is anticipated that this thesis will be completed in May 1990.

4.2.2 Kenneth Krause

Kenneth Krause is presently pursuing research for his doctoral dissertation on the subject of forming images of specular radar targets. A goal is to develop a method that accommodates both diffuse and specular components in the radar echo. It is anticipated that this thesis will be completed in 1991.

4.2.3 Steven Jacobs

Steven Jacobs is presently writing his master's dissertation on the subject of computational issues associated with forming radar images using the estimation methods we have developed. It is anticipated that this thesis will be completed by August 1990.

5. Personnel

The personnel who participated in this research project during the reporting period are the following.

Steven Jacobs

- Graduate Research Assistant in the Electronic Systems and Signals Research Laboratory
- M. Sc. candidate in the Department of Electrical Engineering
- received no support under the O.N.R. Contract
- task: examine computational issues in radar imaging and develop parallel implementation strategies

Kenneth Krause

- Part-Time Graduate Assistant in the Electronic Systems and Signals Research Laboratory
- Employed by the McDonnell-Douglas Astronautics Co.
- Ph. D. Candidate in the Department of Electrical Engineering
- received no support under the O.N.R. Contract
- task: include specular components in the maximum-likelihood method

Pierre A. Moulin

- Graduate Research Assistant in the Electronic Systems and Signals Research Laboratory
- Ph. D. Candidate in the Department of Electrical Engineering
- received support as a Graduate Research Assistant under the O.N.R. Contract
- task: analyze performance of the maximum-likelihood method, include sieve and signal-to-noise ratio constraints for regularization, examine computational issues to make the method practical

Joseph A. O'Sullivan

- Faculty Research Associate in the Electronic Systems and Signals Research Laboratory
- Assistant Professor of Electrical Engineering
- received support as a Senior Research Associate under the O.N.R. Contract
- task: participate in all aspects of the research project

Donald Porter

- Undergraduate Research Assistant in the Electronic Systems and Signals Research Laboratory
- B.S.E.E. and B.S.C.S. degree candidate in the Electrical Engineering and Computer Science Departments
- received no support under the O.N.R. Contract
- task: implement conventional radar imaging algorithms, implement a computer simulation of radar echo data based on a given scattering function

Donald L. Snyder

- Principal Investigator
- Director of the Electronic Systems and Signals Research Laboratory
- Professor of Electrical Engineering
- received support under the O.N.R. Contract
- task: participate in all aspects of the research project

Michael Turmon

- Graduate Research Assistant in the Electronic Systems and Signals Research Laboratory
- M. Sc. candidate in the Department of Electrical Engineering
- received no support under the O.N.R. Contract
- task: study implementation of spectrum-estimation methods of use in radar imaging on a massively parallel computer architecture (1024 element A.M.T. DAP with SUN/4 host)

J. Trent Wohlschlaeger

- Graduate Research Assistant in the Electronic Systems and Signals Research Laboratory
- Ph. D. Candidate in the Department of Electrical Engineering
- received no support under the O.N.R. Contract during the reporting period
- task: study tomographic methods for forming radar images

6. Appendices

6.1 Reprint of: D. Snyder, J. O'Sullivan, and M. Miller, The Use of Maximum Likelihood Estimation for Forming Images of Diffuse Radar Targets from Delay-Doppler Data, *IEEE Trans. on Information Theory*, Vol. 35, pp. 536-548, 1989.

The Use of Maximum Likelihood Estimation for Forming Images of Diffuse Radar Targets from Delay-Doppler Data

Donald L. Snyder
Joseph A. O'Sullivan
Michael I. Miller

Reprinted from
IEEE TRANSACTIONS ON INFORMATION THEORY
Vol. 35, No. 3, May 1989

The Use of Maximum Likelihood Estimation for Forming Images of Diffuse Radar Targets from Delay–Doppler Data

DONALD L. SNYDER, FELLOW, IEEE, JOSEPH A. O'SULLIVAN, MEMBER, IEEE, AND MICHAEL I. MILLER

Abstract—A new approach to high-resolution radar imaging is presented. The starting point is a model of the radar echo signal based on the physics governing radar reflections. This model has been used several times in the past for describing radar targets that are rough compared to the wavelength of the transmitted radiation. Without specifying precisely what the transmitted signal is, a general estimation-based procedure is derived for obtaining images. After discretizing the model, the radar imaging problem reduces to the task of estimating discretized second-order statistics of the reflectance process of the target. Maximum likelihood estimates of these statistics are obtained as the limit point of an expectation-maximization algorithm.

I. INTRODUCTION

RADAR SYSTEMS can be used to produce high-resolution images of a reflecting target. This is accomplished by illuminating the target with a series of pulses and observing the return echoes. Each patch on the target introduces a certain amount of propagation delay and Doppler shift to a pulse it reflects, the amount depending on the range and velocity of the patch relative to the radar transmitter and receiver. The beamwidth of the radar antenna relative to the size of the target is an important consideration. Images can be produced by scanning a narrowly focused beam over the target in some type of raster pattern and then displaying the received power in delay and Doppler or, equivalently, range and cross-range coordinates. Images can also be formed by illuminating the entire target in *spotlight mode* with a wide, relatively unfocused beam. The received signal for each illumination is then a complicated superposition of the echoes received from all the patches that make up the extended surface of the radar target. Our concern will be with forming images of rotating rough targets using a spotlight-mode radar.

We will denote the complex envelope of the signal transmitted by the radar by $(2E_T)^{1/2}s_T(t)$, where E_T is the transmitted energy, and $s_T(t)$ is normalized to unit energy. The particular form of this signal will not need to be specified. The expressions we obtain for producing an image can then be specialized for any signal of choice, including the stepped-frequency and wide-band chirp waveforms used in practice, as discussed by Wehner [1] and Mensa [2] and the chirp-rate modulated waveforms discussed by Bernfeld [3] and Snyder *et al.* [4].[1] When specializing $s_T(t)$, it should be kept in mind that this represents the entire sequence of transmitted pulses that illuminate the target.

Walker [5] gives a clear intuitive description of the radar imaging problem. He considers a small nonfluctuating reflector rotating counterclockwise at the rate of f_r revolutions per second on a circle of radius r centered at a distance R_0 from the radar transmitter/receiver, as shown in Fig. 1. The distance to the reflector at time t is given

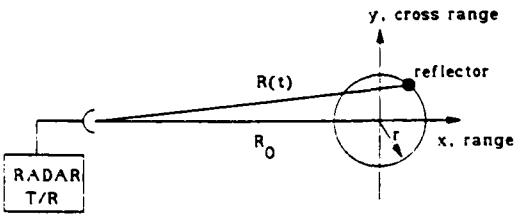


Fig. 1. Geometry for small reflector.

approximately by

$$R(t) \doteq R_0 + x_0\cos(2\pi f_r t) + y_0\sin(2\pi f_r t),$$

provided $R_0 \gg r = (x_0^2 + y_0^2)^{1/2}$, where (x_0, y_0) is the (x, y) position of the reflector at time $t = 0$. Then the radar echo signal $s_R(t)$ received following an illumination by the transmitted signal will be of the form

$$s_R(t) = \sqrt{2E_T}\,s_T(t-\tau)b$$

where $\tau = 2R(t)/c$ is the two-way propagation delay to

Manuscript received June 9, 1987; revised April 1, 1988. This work was supported by the Office of Naval Research under Contract N00014-86-K0370. The material in this paper was partially presented at the 1988 Conference on Information Sciences and Systems, Princeton University, Princeton, NJ, March 1988.

The authors are with the Department of Electrical Engineering, Box 1127, Washington University, St. Louis, MO 63130.

IEEE Log Number 8928182.

[1] *Note added in proof*: It has come to our attention since submitting this manuscript that chirp-rate modulation is also discussed by E. Feig and A. Grünbaum in, "Tomographic methods in range-Doppler radar," *Inverse Probelms*, vol. 2, pp. 185–195, 1986.

0018-9448/89/0500-0536$01.00

the reflector, with c being the propagation velocity. The quantity b is a complex-valued scale factor which models the strength of the received echo. This scale factor will be called the *reflectivity*. It includes the effects of inverse square-law attenuation experienced by the propagating radiation and, importantly, the properties of the reflector that are significant in the electromagnetic scattering interaction, including its shape, size, and surface properties. More generally, the reflectivity can vary with time because the aspect, and therefore, the scattering interaction with the reflector will vary as it rotates. If, as discussed by Walker [5], the radar data $s_R(t)$ are examined over a small interval of time, then the delay τ and Doppler shift f_D can be approximated by

$$\tau \doteq 2c^{-1}(R_0 + x_0)$$

and

$$f_D = \frac{2}{\lambda}\frac{dR}{dt} \doteq \frac{2}{\lambda} y_0 2\pi f_r$$

where λ is the wavelength at the carrier frequency of the radar. Thus the reflectivity b, range x_0, and cross-range y_0, relative to the coordinate axis, can be determined from the amplitude, delay, and Doppler information contained in the radar data. Extracting this information permits the formation of an image of the reflector by displaying $|b|$ or $|b|^2$ at the appropriate location in range and cross-range coordinates. The maximum delay and Doppler shift are determined by the distance $(x_0^2 + y_0^2)^{1/2}$ of the reflector from the coordinate center about which it rotates and the rotation rate f_r; more generally, the extent of a reflector in delay and Doppler is determined by the physical extent of the reflector and the rotation rate.

Now consider a spatially extended target that is rotating. A patch on the surface with a two-way delay in the interval $[\tau, \tau + \Delta\tau)$ reflects a signal that is incident on the patch at time t with a reflectance strength $b(t, \tau)\,\Delta\tau$. Consequently, the complex envelope of the received echo signal $s_R(t)$ following the illumination of the target by $s_T(t)$ is given by the following superposition of returns from reflecting patches at all the two-way delays τ:

$$s_R(t) = \sqrt{2E_T}\int_{-\infty}^{+\infty} s_T(t-\tau)\, b\left(t - \frac{1}{2}\tau, \tau\right) d\tau. \quad (1)$$

The total received signal $r(t)$ is also assumed to be corrupted by an independent additive noise

$$r(t) = s_R(t) + w(t) \quad (2)$$

where $w(t)$ is a complex-valued white Gaussian process with a mean of zero and a covariance function defined by

$$E[w(t)w^*(t')] = N_0\delta(t - t') \quad (3)$$

where the asterisk denotes complex conjugation. We refer to $b(t, \tau)$ as the *reflectance process*. This is a complex-valued random process.

There are two images that may be displayed as the result of processing $r(t)$ with an estimation procedure. One is an estimate of the reflectance process $b(t, \tau)$ itself, and the other is an estimate of the covariance or, equivalently, the spectral density of this process. These may be regarded as conditional first- and second-order statistics of the reflectivity process, respectively, in terms of the radar data (2).

If $b(t, \tau)$ is deterministic, define $c(f, \tau)$ to be its Fourier transform in the t variable,

$$c(f, \tau) = \int_{-\infty}^{+\infty} b(t, \tau) e^{-j2\pi ft}\, dt. \quad (4)$$

The function $c(f, \tau)$ then contains information about the target in delay τ and Doppler f coordinates. An image of the target is obtained by placing the magnitude or squared magnitude of this function into delay and Doppler bins. We refer to this as the *reflectance image*. This transform can be obtained in a variety of ways, depending on the signal $s_T(t)$ selected to illuminate the target. For the stepped-frequency signals used in practice, the usual approach consists of two operations described by Wehner [1]. The first is to place the data into delay (or range) bins by separately Fourier transforming N sample values of the received signal acquired for each transmitted group of N stepped-frequency pulses. The resulting delay-binned data are placed in the rows of an $N \times N$ matrix where each row contains the transformed data from one pulse group. In the second operation, the columns of this matrix are Fourier transformed to obtain a Doppler (or cross-range) profile at each delay. The resulting two-dimensional array is intended to be a discrete version of $c(f, \tau)$ in delay (range) and Doppler (cross-range) coordinates. This processing based on two-dimensional Fourier transforms is derived using a strictly deterministic analysis and so does not account for statistical properties of the reflectance process or for noise that may be present. A similar processing is employed for the linear FM-chirp signals also used in practice for radar imaging [1], [2].

For situations in which the target's surface is rough compared to the wavelength at the carrier frequency, $b(t, \tau)$ may be taken to be a complex-valued Gaussian random process, as discussed by Shapiro *et al.* [6] for radar systems operating at laser frequencies and Van Trees [7] at microwave frequencies. If there are no glint or specular components in the echo, then this is a zero mean process with covariance

$$E[b(t, \tau)b^*(t', \tau')] = K(t - t', \tau)\,\delta(\tau - \tau'). \quad (5)$$

The delta function in this expression results from postulating that each reflecting patch introduces an uncorrelated contribution to the echo signal. That the function $K(t - t', \tau)$ depends only on the difference of t and t', and not on t and t' separately, results from postulating that the reflectance process is wide-sense stationary for each delay. A reflectance process with these properties is said by Van Trees [7] to possess wide-sense stationary uncorrelated scatterers (WSSUS). Assuming that the reflectance process has these properties, the delay–Doppler data associated with the radar target may be obtained from the Fourier transform of $K(t, \tau)$ in the t variable,

$$S(f, \tau) = \int_{-\infty}^{+\infty} K(t, \tau)\exp(-j2\pi ft)\, dt. \quad (6)$$

The function $S(f,\tau)$ is called the *scattering* function of the target and, as a function of f, is the power-density spectrum of the reflectance process at each delay τ. $S(f,\tau)\,\Delta f\Delta\tau$ is the mean-square strength (or power) of the reflectance of all patches on the target having a Doppler shift in the interval $[f, f+\Delta f)$ and a delay in the interval $[\tau, \tau+\Delta\tau)$. The scattering function may be viewed in delay and Doppler coordinates as an image of the target. We call this the *scattering function image*.

Our approach to forming radar images will be to use maximum likelihood methods with the data model in (2) to estimate the scattering function. We will also obtain an estimate of the reflectance process. Model-based approaches that use statistical estimation theory techniques to derive image-formation algorithms appear less frequently in the literature about radar imaging than the deterministic approaches outlined above. One example is that of Frost *et al.* [8], who uses a multiplicative model and Wiener filtering techniques. The approach we will describe differs in that the model (2) we adopt of the echo signal is more complicated than a simple multiplicative one and depends explicitly on the transmitted waveform through a spatial integration over the reflecting target. We also do not restrict the processing to be linear; in particular, we show that the processing of the received data for producing the maximum likelihood estimate of the scattering function and a corresponding estimate of the reflectance process is not linear. An approach for estimating scattering functions of spread channels is given by Gaarder [9], who cites earlier work on the subject by Green [10], Kailath [11], Gallager [12], Hagfors [13], [14], Price [15], Levin [16], Abraham [17], Sifford [18], and Reiffen [19]. Gaarder assumes a specific processing architecture in the form of a cascade of a linear filter square-law envelope detector and another linear filter and claims that this processing is either more general than or equivalent to those of most previous authors. Our approach differs in that no particular processing is assumed in advance; rather, we derive the processing to produce the estimates, starting from a model for the received data and applying recent results in maximum likelihood estimation. The processing which results is quite distinct from that discussed by Gaarder.

For our new approach to radar imaging, we adopt the WSSUS model of a diffuse radar target described by Shapiro *et al.* [6] and Van Trees [7]. We treat both the reflectance process and its second-order statistic, the scattering function, as unknown quantities. The iterative approach we develop for forming images yields the maximum-likelihood estimate of the scattering function and, simultaneously, the conditional-mean estimate of the reflectance process based on statistics which are consistent with the estimated scattering function. Thus both of the quantities treated separately in other imaging schemes are produced simultaneously with our new approach, which is a unique and important aspect of our approach.

We will develop a necessary condition, called the *trace condition*, which the maximum likelihood estimate of the target's scattering function must satisfy. This equation appears to be very hard to solve analytically. As a consequence, we reformulate the imaging problem using the concept of incomplete–complete data spaces and then use the expectation-maximization algorithm of Dempster *et al.* [20] to derive an iterative algorithm for producing the maximum likelihood estimate of the scattering function. The technique we use to accomplish this parallels that described by Miller and Snyder [21] for power-spectrum estimation and extends their work to include indirect measurements of the process whose spectrum is sought; the process is now measured following the linear transformation and additive noise seen in (2). As shown by Turmon and Miller [22], this is a high-resolution approach to spectrum estimation which results in estimated spectra with smaller bias and mean-square error than other recently developed approaches discussed in the literature. We expect that similar improvements will be seen in radar images of scintillating, diffuse targets when this new technique is used.

II. Maximum Likelihood Imaging for the Incomplete Data Model

For reasons that will become evident in the next section, we term the data $r(t)$ in (2) the *incomplete* data for the radar-imaging problem. The model given in the Introduction for these data consists of the sum of the radar echo signal $s_R(t)$ of (1) and an independent white noise process $w(t)$. We may, therefore, state the problem of imaging a diffuse radar target as that of estimating the scattering function $S(f,\tau)$ or equivalently, the covariance function $K(t,\tau)$ given radar-return data $\{r(t),\ T_i \le t \le T_f\}$ on an observation interval (T_i, T_f). In this section, we first discretize the model for the incomplete data and then derive and discuss a necessary condition, called the *trace condition*, which the maximum likelihood estimate of the discretized version of $K(t,\tau)$ must satisfy.

Discrete Model

In anticipation of using discrete-time processing of radar data to produce images, we now state the discrete version of our model as follows. We are given N samples of the complex-valued radar data corresponding to (2),

$$r(n) = s_R(n) + w(n), \qquad n = 0,1,\cdots,N-1 \tag{7}$$

where $w(n)$ is a white Gaussian sequence with zero mean and covariance

$$E[w(n)w^*(n')] = N_0\delta_{n,n'} \tag{8}$$

where $\delta_{n,n'}$ is the Kronecker delta function, and the signal samples corresponding to (1) are given by

$$s_R(n) = \sqrt{2E_T}\sum_{i=-\infty}^{+\infty} s_T(n,i)b(n,i),$$
$$n = 0,1,\cdots,N-1. \tag{9}$$

In this expression, we define $s_T(n,i)$ and $b(n,i)$ in terms of the transmitted signal and the reflectance process, re-

spectively, according to

$$s_T(n,i) = s_T(n\,\Delta t - i\,\Delta\tau) \tag{10}$$

and

$$b(n,i) = b\left(n\,\Delta t - \frac{1}{2}i\,\Delta\tau, i\,\Delta\tau\right)\Delta\tau \tag{11}$$

where Δt and $\Delta\tau$ are the sampling intervals adopted in the discretization in time and delay, respectively. We assume that the target has a finite extent in range; thus $b(n,i)$ and therefore terms forming the sum in (9) are zero for i outside the I_R (here, the subscript R denotes range) values $m, m+1, \cdots, m+I_R-1$ starting from the minimum two-way delay corresponding to m. This discrete reflectance is a Gaussian sequence with zero mean and covariance given by

$$E[b(n,i)b^*(n',i')] = K(n-n',i)\,\delta_{i,i'}. \tag{12}$$

The discrete scattering function $S(v,i)$ is the Fourier transform of $K(n,i)$ in the n variable,

$$S(v,i) = \sum_{n=-\infty}^{+\infty} K(n,i)\exp(-j2\pi vn). \tag{13}$$

The imaging problem for the discrete model is to estimate $S(v,i)$ or, equivalently, the covariance function $K(n,i)$, for all frequencies v spanning the target in Doppler, and for all delays i spanning the target in the delay, given the radar data $\{r(n),\ n=0,1,\cdots,N-1\}$.

Matrix Model

These discrete equations may conveniently be written in matrix form as follows. Define r to be the received-signal vector of dimension N,

$$r = \begin{pmatrix} r(0) \\ r(1) \\ \vdots \\ r(N-1) \end{pmatrix} = s_R + w \tag{14}$$

where the N-dimensional vectors s_R and w are given by

$$s_R = \begin{pmatrix} s_R(0) \\ s_R(1) \\ \vdots \\ s_R(N-1) \end{pmatrix} \qquad w = \begin{pmatrix} w(0) \\ w(1) \\ \vdots \\ w(N-1) \end{pmatrix}. \tag{15}$$

Also, define S^* as the $NI_R \times N$ rectangular matrix expressed in column-block form in terms of I_R separate $N \times N$ matrices according to

$$S^* = \begin{pmatrix} S_0 \\ S_1 \\ \vdots \\ S_{I_R-1} \end{pmatrix} \tag{16}$$

where S_j is an $N \times N$ diagonal matrix containing sample values of the complex envelope of the transmitted signal $s_T(t)$,

$$S_j = \begin{pmatrix} s_T(0,m+j) & 0 & 0 & \cdots & 0 \\ 0 & s_T(1,m+j) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & & 0 \\ \vdots & \vdots & \cdots & & \vdots \\ 0 & 0 & \cdots & & s_T(N-1,m+j) \end{pmatrix}. \tag{17}$$

Further, define the reflectance vector b of dimension NI_R in the column-block form of I_R vectors according to

$$b = \begin{pmatrix} b(0) \\ b(1) \\ \vdots \\ b(I_R-1) \end{pmatrix} \tag{18}$$

where each $b(i)$ is a vector of dimension N,

$$b(i) = \begin{pmatrix} b(0,m+i) \\ b(1,m+i) \\ \vdots \\ b(N-1,m+i) \end{pmatrix}. \tag{19}$$

Using (9) and these definitions, we can now express the N-dimensional signal vector s_R of (14) and (15) as

$$s_R = \sqrt{2E_T}\,S^+ b \tag{20}$$

where a superscript plus sign denotes the Hermitian-transpose operation. In terms of these defined matrices, the received vector has zero mean and covariance

$$\begin{aligned} K_r &= E(rr^+) = E(s_R s_R^+) + E(ww^+) \\ &= 2E_T S^+ E(bb^+) S + N_0 I. \end{aligned} \tag{21}$$

Then, since

$$E(b(i)b^+(j)) = K(i)\,\delta_{i,j} \tag{22}$$

where $K(i)$ is the Hermitian-symmetric Toeplitz matrix

$$K(i) = \begin{pmatrix} K(0,m+i) & K^*(1,m+i) & \cdots & K^*(N-1,m+i) \\ K(1,m+i) & K(0,m+i) & \cdots & \vdots \\ \cdot & \vdots & \cdots & \\ K(N-1,m+i) & & \cdots & K(0,m+i) \end{pmatrix}, \tag{23}$$

it follows from (21) that the covariance K_r of r is given by

$$K_r = 2E_T S^+ KS + N_0 I \tag{24}$$

where K is the block-diagonal $NI_R \times NI_R$-dimensional matrix defined by

$$K = \begin{pmatrix} K(0) & 0 & 0 & \cdots & 0 \\ 0 & K(1) & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & K(I_R-1) \end{pmatrix}. \tag{25}$$

The ith diagonal block $K(i)$ of K is the covariance matrix of the reflectance process at the ith delay bin.

The Estimation Problem

We will use the following definition.

Definition: Let $\boldsymbol{K}$ denote the set of all $NI_R \times NI_R$ block-diagonal matrices (25) with each block $K(i)$ an $N \times N$ Hermitian-symmetric Topelitz matrix (23). Let $\Omega \subset \boldsymbol{K}$ be a specified convex subset of $\boldsymbol{K}$. Any matrix $K \in \Omega$ is termed admissible. A variational matrix $\delta K \in \boldsymbol{K}$ is called an admissible variation of K for a fixed $K \in \Omega$ if there exists an $\alpha > 0$ such that $K + \beta \delta K \in \Omega$ for all β satisfying $0 < \beta < \alpha$.

The problem is to form an admissible estimate of the covariance matrix K of (25) given the data vector r of (14). The radar image then viewed is the discrete scattering function, an estimate of which may be obtained from the estimate of K by use of (13).

In the foregoing definition the constraint that K be in Ω is used to obtain a "reasonable" setup of the problem. Here we assume that the scattering function $S(f, \tau)$ in (6) is only nonzero for frequencies f satisfying $|f| \le f_{\max}$ for some finite upper frequency $f_{\max}$ and for all delays i. This is equivalent to the assumption of a target of finite cross section and rotation rate. The discrete-time scattering function $S(v, i)$ of (13) is then a periodic function of v consisting of a sum of shifted replicas of $S(f, \tau)$ scaled in amplitude by $1/\Delta t$ and in frequency by Δt, where Δt is the time between samples of $r(t)$. The replicas of $S(f, \tau)$ are centered at every integer on the v scale. To guarantee that there is no aliasing, assume that the sample rate $1/\Delta t$ of $r(t)$ satisfies the Nyquist condition $1/\Delta t > 2f_{\max}$. Then $S(v, i)$ will be nonzero between $-1/2$ and $+1/2$ only for v satisfying $|v| \le v_{\max} = f_{\max} \Delta t$. The output of our algorithm is $S(v, i)$ discretized in frequency v. For a resolution having at least I_{CR} (here, the subscript CR denotes cross range) samples in the frequency range $-v_{\max} \le v \le v_{\max}$, a total of

$$P > \frac{I_{CR}}{2\Delta t f_{\max}}$$

samples of v between $-1/2$ and $+1/2$ are required.

The model of the incomplete data r of (14) is that r is normally distributed with zero mean and covariance specified in (24). Given the incomplete data, we wish to estimate the covariance K of the reflectance process, as defined in (25). To do this, we adopt the maximum likelihood method of statistics, which selects K to maximize the incomplete data log likelihood

$$L_{id}(K) = -\ln\left(\det\left(2E_T S^+ KS + N_0 I\right)\right) - r^+ \left(2E_T S^+ KS + N_0 I\right)^{-1} r \tag{26}$$

where the maximization is subject to the constraint that K be an admissible matrix.

Lemma 1: A necessary condition for an admissible K in the interior of Ω to be a local maximum of $L_{id}(K)$ over all $K \in \Omega$ is

$$\operatorname{tr}\left(\left(2E_T S^+ KS + N_0 I\right)^{-1}\left(rr^+ - 2E_T S^+ KS - N_0 I\right) \cdot \left(2E_T S^+ KS + N_0 I\right)^{-1} S^+ \delta KS\right) = 0 \tag{27}$$

for all admissible variations δK.

The proof of Lemma 1 in the Appendix is based on the fact that the necessary condition for an admissible K to maximize $L_{id}(K)$ is that, for all admissible variations δK,

$$\lim_{\alpha \downarrow 0^+} \frac{L_{id}(K + \alpha \delta K) - L_{id}(K)}{\alpha} \le 0. \tag{28}$$

We call (27) the *trace condition*. Burg *et al.* [23] have studied an equivalent problem of Toeplitz-constrained covariance estimation and have derived the trace condition using a different approach.

If $\Omega = \boldsymbol{K}$, there are NI_R unknowns in K. Since $\delta K \in \boldsymbol{K}$, there are NI_R parameters in δK that can be varied for this case. These variations in the trace condition generate NI_R equations in the unknown elements of K. Thus, in principle, the trace condition produces enough equations to determine the unconstrained maximum likelihood estimate K. However, the equations are complicated due to the inverse matrices appearing in (27); thus it does not appear feasible to determine K directly from the trace condition. This motivates the development of the iterative approach presented in the next section. A sequence of estimates that increase the likelihood at each iteration stage is identified, and we demonstrate that stable points of the iteration satisfy the trace condition.

The trace condition is only a necessary condition that the estimate K must satisfy. For it to be sufficient as well, the second derivative must be negative along all admissible variational directions δK.

Lemma 2: Sufficient conditions for an admissible matrix K in the interior of Ω to be a local maximum of $L_{id}(K)$ are that, first, the trace condition (27) is satisfied for all admissible variations δK and, second, that

$$\operatorname{tr}\left(K_r^{-1} S^+ \delta KS K_r^{-1}\left(2E_T S^+ KS + N_0 I - 2rr^+\right) K_r^{-1} S^+ \delta KS\right) < 0 \tag{29}$$

for all admissible variations δK.

The proof of Lemma 2 is given in the Appendix. Equation (29) is just the second derivative of $L_{id}(K)$ in the direction δK.

III. Maximum Likelihood Imaging for the Incomplete/Complete Data Model

The fact that the trace condition (27) cannot be solved directly for the maximum likelihood estimate of K motivates the indirect approach we now take of embedding the imaging problem in a larger, seemingly more difficult problem. The result will be an iterative algorithm which, when implemented, produces a sequence of admissible matrices $K^{(0)}, K^{(1)}, \cdots, K^{(k)}, \cdots$ with the property that the corresponding sequence of incomplete data log likelihoods $L_{id}[K^{(0)}]$, $L_{id}[K^{(1)}], \cdots, L_{id}[K^{(k)}], \cdots$ is nondecreasing at each stage.

Fuhrmann and Miller [24] have recently shown that maximum likelihood estimates of Toeplitz-constrained covariances which are positive definite do not always exist when given only one data vector r. A necessary and sufficient condition for the likelihood function to be unbounded, and therefore for no maximum likelihood estimate to exist, is that there be a singular Toeplitz matrix with the data in its range space. For our imaging problem, this condition is that an admissible K exists with

$$2E_T S^+ KS + N_0 I$$

singular so that

$$r = (2E_T S^+ KS + N_0 I)\alpha \tag{30}$$

for some complex-valued vector α. In fact, with only a Toeplitz constraint on K, a sufficient condition that a singular estimate for K be obtained is that $N_0 = 0$ and that a singular K exist with r in the range space of $2E_T S^+ KS$. The argument for this mirrors that of Fuhrmann and Miller in [24, theorem 1] but is applied to the complete data log likelihood given in (A7) of the Appendix. Fuhrmann and Miller also showed that, even if the true covariance had eigenvalues bounded from above and below, the probability that a singular Toeplitz matrix exists with the data in its range can be very close to one. By restricting the search to Toeplitz matrices with circulant extensions, they were able to show that the probability a singular circulant Toeplitz matrix has the data in its range space is zero. Thus, for maximum likelihood estimates to be nonsingular with probability one for all nonnegative values of N_0, we may restrict the class of admissible Toeplitz matrices to be those with circulant extensions of period P, where P is equal to or greater than the number N of data samples available $P > N$. This is not a severe restriction because the set of all Toeplitz matrices is approached by the subset having circulant extensions of period P as P tends to infinity. What we envision in adopting this constraint is that for each delay i, the N sample values of the reflectance $b(n, i)$, $n = 0, 1, \cdots, N-1$, are from a stationary process that is periodic with period P, where P could be some large but finite value; a lower bound on P in terms of a desired cross-range resolution is discussed above. These N sample values of the reflectance enter the incomplete data r according to (14) and (20). By using the expectation-maximization algorithm of Dempster *et al.* [20], we shall develop a sequence of admissible matrices that have the maximum likelihood estimate of K subject to this circulant extension as a stable point. The approach parallels that of Miller and Snyder [21] for estimating the power spectrum of a time series from a single set of data. An important benefit of introducing the periodic extension and using the expectation-maximization algorithm is that estimates of both the scattering function and the reflectance processes are obtained simultaneously and can be readily viewed as target images in range and cross-range coordinates; thus the procedure proposed may be considered natural for the imaging problem because both types of images considered separately in the past are obtained directly. As a final comment regarding our use of a circulant extension for K, we note that in estimating a discretized version of the target's scattering function, the class of admissible K is restricted automatically to consist of those with circulant extensions. For completeness, we also include in the Appendix the equations obtained using the expectation-maximization algorithm for estimating general Toeplitz matrices when the assumption of a circulant extension is not made.

We shall introduce a modification of our notation to indicate that the N samples of the reflectance process are from a stationary periodic process of period P. To this end let $b_N(i)$ denote the N-dimensional vector $b(i)$ of (19). We now think of $b_N(i)$ as an N-dimensional subvector of the P-dimensional vector $b_P(i)$ formed from samples of the reflectance process over a full period,

$$b_P(i) = \begin{pmatrix} b(0, m+i) \\ b(1, m+i) \\ \vdots \\ b(N-1, m+i) \\ \vdots \\ b(P-1, m+i) \end{pmatrix}. \tag{31}$$

If I_N is the $N \times N$ identity matrix, and J_R the $P \times N$ matrix defined by

$$J_R = \begin{pmatrix} I_N \\ 0 \end{pmatrix}, \tag{32}$$

then

$$b_N(i) = J_R^+ b_P(i).$$

Also, let b_N denote the NI-dimensional vector b of (18), and b_P the PI-dimensional vector with ith block element $b_P(i)$. Then,

$$b_N = M_R^+ b_P$$

where M_R is the $PI_R \times NI_R$ block-diagonal matrix

$$M_R = \begin{pmatrix} J_R & 0 & 0 & \cdot & \cdot & 0 \\ 0 & J_R & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & J_R \end{pmatrix}. \tag{33}$$

Furthermore, let $K_N(i)$ denote the $N \times N$ Toeplitz covariance matrix $K(i)$ of $b_N(i)$ defined in (23), and $K_P(i)$ the $P \times P$ circulant covariance matrix of $b_P(i)$. Then, the

Toeplitz matrix $K_N(i)$ is the upper left block of the circulant matrix $K_P(i)$, as given by

$$K_N(i) = J_R^+ K_P(i) J_R.$$

Lastly, let K_P denote the $PI_R \times PI_R$ block-diagonal matrix in the form of (25) with the ith diagonal block being $K_P(i)$. Then, if K_N denotes the $NI_R \times NI_R$ matrix K of (25),

$$K_N = M_R^+ K_P M_R.$$

Let W denote the $P \times P$ discrete Fourier-transform matrix scaled so that the columns are orthonormal,

$$W = \frac{1}{\sqrt{P}} \begin{pmatrix} w_P^0 & w_P^0 & \cdot & \cdot & w_P^0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ w_P^0 & w_P^k & w_P^{2k} & \cdot & w_P^{(P-1)k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ w_P^0 & w_P^{P-1} & \cdot & \cdot & w_P^{(P-1)(P-1)} \end{pmatrix} \quad (34)$$

where $w_P = \exp(-j2\pi/P)$. Also, let W_P be the $PI_R \times PI_R$ block-diagonal matrix

$$W_P = \begin{pmatrix} W & 0 & 0 & \cdot & \cdot & 0 \\ 0 & W & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & W \end{pmatrix}. \quad (35)$$

Then, b_P can be represented in rotated coordinates according to

$$a_P = W_P b_P = \begin{pmatrix} a(0) \\ a(1) \\ \vdots \\ a(I_R - 1) \end{pmatrix} \quad (36)$$

where $a(i) = Wb_P(i)$. The assumption that $b_P(i)$ originates from a periodic process implies that the PI_R-dimensional vector a_P is normally distributed with zero mean and diagonalized covariance

$$A_P = \mathrm{E}(a_P a_P^+) = W_P K_P W_P^+. \quad (37)$$

We will denote the $(p + iP)$th diagonal element of A_P by $\sigma_p^2(i)$; this is the pth diagonal element of the $P \times P$ diagonal matrix $\mathrm{E}[a(i)a^+(i)]$.

Let $S(v, i)$ be discretized in frequency with P samples taken for $0 \le v < 1$. These samples may be obtained from (13) as

$$S\left(\frac{p}{P}, i\right) = \sum_{n=0}^{P-1} K_P(n, i) \exp\left(-j\frac{2\pi np}{P}\right) \quad (38)$$

for $p = 0, 1, \cdots, P-1$. The pth such sample is just the pth entry in the vector

$$\sqrt{P}\, W K_P(i) e \quad (39)$$

where e is the P-dimensional unit vector

$$e = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Substituting (37) into (39), we get

$$\sqrt{P}\, W K_P(i) e = \sqrt{P}\, A(i) W e, \quad (40)$$

but $P^{1/2} We$ is a P-dimensional vector of ones, and therefore

$$S\left(\frac{p}{P}, i\right) = \sigma_p^2(i), \quad (41)$$

which, according to the above definition, is the $(p + iP)$th diagonal element of the diagonal matrix A_P. The entries of the diagonal matrix A_P in (37) are then samples of the scattering function.

The constraint from Section II that the scattering function $S(v,i)$ be nonzero only for $|v| \le f_{\max} \Delta t = v_{\max}$, for values of v between $-1/2$ and $+1/2$, may be incorporated at this point in the development. Since A_P is a diagonal matrix of samples of the scattering function, we restrict A_P to have nonzero values only in its top left and bottom right corners. More precisely, let I_{CR} be the smallest odd integer satisfying

$$I_{CR} > 2v_{\max} P.$$

I_{CR} is the number of cross-range resolution cells implied by P and $v_{\max}$. Then, let J_{CR} be the $I_{CR} \times P$ matrix

$$J_{CR} = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & 0 & I_2 \end{pmatrix}$$

where I_1 is an $[(I_{CR}+1)/2] \times [(I_{CR}+1)/2]$ identity matrix and I_2 is an $[(I_{CR}-1)/2] \times [(I_{CR}-1)/2]$ identity matrix. Let M_{CR} be the $I_{CR} I_R \times PI_R$ matrix

$$M_{CR} = \begin{pmatrix} J_{CR} & 0 & 0 & \cdot & \cdot & 0 \\ 0 & J_{CR} & 0 & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & J_{CR} \end{pmatrix}.$$

Define Σ_P to be the diagonal matrix

$$\Sigma_P = M_{CR} A_P M_{CR}^+.$$

The diagonal elements of Σ_P are the potentially nonzero diagonal elements of A_P. Recognizing that some elements of the diagonal matrix A_P are zero and using the definition of M_{CR}, we conclude also that

$$A_P = M_{CR}^+ \Sigma_P M_{CR}.$$

The set Ω referred to in the definition in Section II can now be specified. We restrict consideration to those covariance matrices generated by Σ_P from above, so

$$\Omega = \{ K \in K : K = M_R^+ W_P^+ M_{CR}^+ \Sigma_P M_{CR} W_P M_R \}.$$

For use with the expectation-maximization algorithm, we identify the complete data as (c_P, w), where w is the N-dimensional noise vector defined in (15) and c_P is defined by

$$c_P = M_{CR} a_P.$$

Since elements of a_P corresponding to the zero diagonal elements of A_P are almost surely zero, we see also that

$$a_P = M_{CR}^+ c_P.$$

Using this fact, we note from (14) and (20) that the incomplete data r can be obtained from the complete data according to the mapping

$$r = \Gamma^{+} c_P + w \tag{42}$$

where we define the $I_R I_{CR} \times N$ matrix (which appears throughout the development that follows):

$$\Gamma = \sqrt{2E_T} M_{CR} W_P M_R S.$$

The log likelihood function $L_{cd}(\Sigma_P)$ of the complete data as a function of Σ_P, the diagonal covariance matrix of c_P, is given by

$$\begin{aligned} L_{cd}(\Sigma_P) &= -\ln(\det(\Sigma_P)) - c_P^{+} \Sigma_P^{-1} c_P \\ &= -2 \sum_{i=0}^{I_R-1} \sum_{p=0}^{I_{CR}-1} \ln(\sigma_p(i)) \\ &\quad - \sum_{i=0}^{I_R-1} \sum_{p=0}^{I_{CR}-1} |c_p(i)|^2 \sigma_p^{-2}(i) \end{aligned} \tag{43}$$

where all terms that are not a function of Σ_P have been suppressed and $c_p(i)$ is the pth element of the I_{CR}-dimensional vector $c_P(i) = J_{CR} a(i)$.

The expectation-maximization algorithm for estimating the covariance of the reflectance process K_P from the incomplete data r is an alternating maximization procedure in which a sequence of estimates $\Sigma_P^{(0)}, \Sigma_P^{(1)}, \cdots, \Sigma_P^{(k)}, \cdots$ of Σ_P is obtained first, where the expectation-maximization procedure specifies how to obtain $\Sigma_P^{(k+1)}$ from $\Sigma_P^{(k)}$ for $k = 0, 1, \cdots$. If $\Sigma_P^{(k)}$ denotes the estimate of Σ_P at stage k, then there is a corresponding term

$$K_P^{(k)} = W_P^{+} M_{CR}^{+} \Sigma_P^{(k)} M_{CR} W_P$$

in a sequence of estimates of K_P. Likewise, to the kth term $K_P^{(k)}$ of the sequence of estimates of K_P, there is a corresponding term

$$K_N^{(k)} = M_R^{+} K_P^{(k)} M_R$$

in a sequence of estimates of K_N. These have increasing log likelihoods $L_{id}[K_N^{(0)}] \le L_{id}[K_N^{(1)}] \le \cdots \le L_{id}[K_N^{(k)}] \le \cdots$. We show that stable points of this sequence satisfy the necessary trace condition for the maximum likelihood estimate of K_N, where $K_N^{(i)}$ is a stable point if $K_N^{(i+j)} = K_N^{(i)}$ for $j = 1, 2, \cdots$.

Each iteration stage of the expectation-maximization algorithm has an expectation E step and a maximization M step that must be performed to get to the next step. The E-step requires evaluation of the conditional expectation of the complete data log likelihood (43) given the incomplete data r and assuming that the covariance defining the complete data is $\Sigma_p^{(k)}$,

$$Q\left[\Sigma_P | \Sigma_P^{(k)}\right] = \mathrm{E}\left[L_{cd}(\Sigma_P) | r, \Sigma_P^{(k)}\right]. \tag{44}$$

Since L_{cd} in (43) is a function of the nonrandom matrix Σ_P, the result of this conditional expectation is a function of Σ_P. It is also a function of $\Sigma_P^{(k)}$ because the expectation is performed assuming that c_p is normally distributed with covariance $\Sigma_P^{(k)}$. We have indicated this dependency of the conditional expectation on both Σ_P and $\Sigma_P^{(k)}$ in the definition of the function Q in (44). From (43), we have that

$$\begin{aligned} Q\left[\Sigma_P | \Sigma_P^{(k)}\right] &= -2 \sum_{i=0}^{I_R-1} \sum_{p=0}^{I_{CR}-1} \ln(\sigma_p(i)) \\ &\quad - \sum_{i=0}^{I_R-1} \sum_{p=0}^{I_{CR}-1} \mathrm{E}\left[|c_p(i)|^2 | r, \Sigma_P^{(k)}\right] \sigma_p^{-2}(i). \end{aligned} \tag{45}$$

The M-step yields the estimate $\Sigma_P^{(k+1)}$ at stage $k+1$ as the choice of Σ_P that maximizes this conditional expectation,

$$\Sigma_P^{(k+1)} = \arg\max\left[Q\left(\Sigma_P | \Sigma_P^{(k)}\right)\right], \tag{46}$$

subject to the constraint that the maximizer be a diagonal covariance matrix. From (45), this maximization yields the diagonal matrix $\Sigma_P^{(k+1)}$ with $(p + iI_{CR})$th diagonal element

$$\left(\sigma_p^2(i)\right)^{(k+1)} = \mathrm{E}\left[|c_p(i)|^2 | r, \Sigma_P^{(k)}\right]. \tag{47}$$

Thus, we may write $\Sigma_P^{(k+1)}$ as

$$\Sigma_P^{(k+1)} \overset{d}{=} \mathrm{E}\left[c_P c_P^{+} | r, \Sigma_P^{(k)}\right] \tag{48}$$

where the d over the equal sign means that the diagonal terms in the matrix on the left side equal the diagonal terms in the matrix on the right side and that all the off diagonal elements on the left side are zero.

Expression (48) appears to be complicated because of the several matrices we have defined, but it produces a sequence of covariance estimates having a straightforward interpretation. If we form the matrix $K_P^{(k+1)}$ according to

$$K_P^{(k+1)} = W_P^{+} M_{CR}^{+} \Sigma_P^{(k+1)} M_{CR} W_P, \tag{49}$$

we then find that

$$K_P^{(k+1)} = \begin{pmatrix} K_P^{(k+1)}(0) & 0 & 0 & \cdots & 0 \\ 0 & K_P^{(k+1)}(1) & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & K_P^{(k+1)}(I_R - 1) \end{pmatrix} \tag{50}$$

where $K_P^{(k+1)}(i)$ is a $P \times P$ circulant matrix interpreted as the estimate at stage $k+1$ of the covariance $K_P(i)$ of the P-periodic reflectance process at delay $m+i$. Miller and Snyder [21] show that the (n, m)-element of this circulant matrix is given by

$$\frac{1}{P} \sum_{p=0}^{P-1} \mathrm{E}\left[b(p, i) b^{*}(\langle p + m - n \rangle_P, i) | r, K_P^{(k)}\right] \tag{51}$$

where $\langle a \rangle_P = a \bmod P$. Equation (51) has an intuitively appealing form. If the reflectivity process $b(n, i)$ could be observed for all instants $n = 0, 1, \cdots, P-1$ in a period and for each i independently, then the maximum likelihood estimate of the covariance $K_P(i)$ would be the arithmetic

average of the lagged products

$$\frac{1}{P}\sum_{p=0}^{P-1} b(p,i)b^*(\langle p+m-n\rangle_P, i). \tag{52}$$

Equations (50) and (51) indicate that one should simply substitute the conditional mean estimate of an unknown lagged product into this expression to form the maximum likelihood estimate of the covariance when only the incomplete data are known.

Estimating Σ_p and K_p

The maximum likelihood estimate of Σ_p is a stable point of the sequence defined in (48). The terms on the right side of this equation can be evaluated as follows. Let the conditional-mean estimate of c_p in terms of the incomplete data r be defined at stage k by

$$\hat{c}_p^{(k)} = \mathrm{E}\left[c_p | r, \Sigma_p^{(k)}\right]. \tag{53}$$

Then, (48) can be rewritten in the form

$$\Sigma_p^{(k+1)} \stackrel{d}{=} \mathrm{E}\left[\left(c_p - \hat{c}_p^{(k)}\right)\left(c_p - \hat{c}_p^{(k)}\right)^+ | r, \Sigma_p^{(k)}\right] + \hat{c}_p^{(k)}\hat{c}_p^{(k)+}. \tag{54}$$

Now examination of (42) shows that forming the conditional-mean estimate (53) of c_p from r is a standard problem in linear estimation theory. From Tretter [25, ch. 14], for example, we find that

$$\hat{c}_p^{(k)} = \Sigma_p^{(k)}\Gamma\left[\Gamma^+\Sigma_p^{(k)}\Gamma + N_0 I\right]^{-1} r. \tag{55}$$

Furthermore, the first term on the right side in (54) is the covariance of the estimation error when c_p is estimated from r. Also from Tretter [25, ch. 14], we have

$$\mathrm{E}\left[\left(c_p - \hat{c}_p^{(k)}\right)\left(c_p - \hat{c}_p^{(k)}\right)^+ | r, \Sigma_p^{(k)}\right] = \Sigma_p^{(k)} - \Sigma_p^{(k)}\Gamma\left[\Gamma^+\Sigma_p^{(k)}\Gamma + N_0 I\right]^{-1}\Gamma^+\Sigma_p^{(k)}. \tag{56}$$

In summary, the following steps are performed to produce a sequence $\Sigma_p^{(0)}, \Sigma_p^{(1)}, \cdots, \Sigma_p^{(k)}, \cdots$ of estimates of Σ_p fo which the corresponding sequence of likelihoods is nondecreasing:

1) set $k = 0$, selecting a starting estimate $\Sigma_p^{(0)}$;
2) calculate the estimate of c_p according to (55);
3) calculate the error covariance according to (56);
4) update the estimate of Σ_p according to (54);
5) if "last iteration" then stop, else replace k by $k+1$ and go to 2.

The starting value in step 1 can be any positive-definite diagonal covariance matrix of dimension $I_R I_{CR} \times I_R I_{CR}$. A sequence of estimates of K_p having increasing likelihood is obtained from the sequence of estimates of Σ_p according to (49).

Forming the Scattering-Function Image

From (41), the diagonal elements of the $I_R I_{CR} \times I_R I_{CR}$-dimensional matrix Σ_p are sample values of the scattering function, with the scattering function samples at delay $m+i$ given by the I_{CR} diagonal elements of the ith $I_{CR} \times I_{CR}$-dimensional diagonal block of Σ_p. We may, therefore, simply regard $\Sigma_p^{(k)}$ as the stage k estimate of the scattering function. The stage-k scattering-function image of the target in range (i coordinate) and cross range (p coordinate) can be displayed as follows. Let $\Sigma_p^{(k)}(i)$ denote the ith $I_{CR} \times I_{CR}$-dimensional diagonal block of $\Sigma_p^{(k)}$, and denote the pth diagonal element of $\Sigma_p^{(k)}(i)$ by $s(p,i) = [\sigma_p^2(i)]^{(k)}$ for $p = 0, 1 \cdots, I_{CR}-1$. Then, $s(0,i)$ is displayed at range $m+i$ and cross range corresponding to a Doppler shift of zero; $s(1,i)$ and $s(I_{CR}-1,i)$ are displayed at range $m+i$ and cross range corresponding to a Doppler shift of $v = 1/P$ and $v = -1/P$, respectively; $s(2,i)$ and $s(I_{CR}-2,i)$ are displayed at range $m+i$ and cross range corresponding to a Doppler shift of $v = 2/P$ and $v = -2/P$, respectively; and so forth, with $s(p,i)$ and $s(I_{CR}-p,i)$ displayed at range $m+i$ and cross range corresponding to a Doppler shift of $\pm p/P$ for $p = 1, 2, \cdots, (I_{CR}-1)/2$ when I_{CR} is odd.

Forming the Reflectance-Process Image

It is interesting to note that the kth stage conditional-mean estimate of c_p, given the measurements r and assuming that the second-order statistics of reflectance are given by the kth stage estimate of the scattering function, is used to form the estimate of Σ_p at stage $k+1$ when the expectation-maximization algorithm is used. This estimate is of very much interest in its own right because, from (36) and its definition, the I_{CR} elements of $c_p(i)$ are the potentially nonzero Fourier-transform coefficients of the reflectance process $b_p(i)$. The target's reflectance image at stage k is formed by placing these elements at range $m+i$ and cross range in the same manner as described above for the scattering-function image.

Convergence Issues

There are some important properties of the iteration sequence (48) which are worth mentioning. First, each step is in an improving direction in the sense that the log likelihood increases at every step and continues to do so until a stable point is reached. This is shown by writing (54) out as

$$\Sigma_p^{(k+1)} \stackrel{d}{=} \Sigma_p^{(k)} + \Sigma_p^{(k)}\Theta^{(k)}\Sigma_p^{(k)} \tag{57}$$

where

$$\Theta^{(k)} = \Gamma K_r^{(k)-1}\left(rr^+ - K_r^{(k)}\right)K_r^{(k)-1}\Gamma^+ \tag{58}$$

and where

$$K_r^{(k)} = \Gamma^+\Sigma_p^{(k)}\Gamma + N_0 I \tag{59}$$

is the kth estimate of the covariance K_r of r. Next, the trace condition (27) which the maximum likelihood estimate must satisfy is reexamined. From the assumption of the P-periodicity of the reflectance process and the matrix definitions given, the admissible variations δK must be of the form

$$\delta K = M_R^+ W_P^+ M_{CR}^+ \delta\Sigma\, M_{CR} W_P M_R. \tag{60}$$

Here $\delta\Sigma$ is a diagonal matrix of the same dimensions as Σ. The trace condition (27) then becomes

$$(2E_T)^{-1}\mathrm{tr}\left(K_r^{-1}(rr^+ - \Gamma^+\Sigma_p\Gamma - N_0 I)K_r^{-1}\Gamma^+\delta\Sigma\,\Gamma\right) = 0. \tag{61}$$

Using the fact that $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ and evaluating this trace at the kth iterate, we see that the trace on the left side of (61) is equal to

$$(2E_T)^{-1}\mathrm{tr}\left(\Theta^{(k)}\delta\Sigma\right). \tag{62}$$

According to (57), $\Sigma_p^{(k)}$ is changed at each stage by adding the diagonal elements of

$$\Sigma_p^{(k)}\Theta^{(k)}\Sigma_p^{(k)} \tag{63}$$

to $\Sigma_p^{(k)}$. Define

$$\overline{\delta\Sigma}^{(k)} \stackrel{d}{=} \Sigma_p^{(k)}\Theta^{(k)}\Sigma_p^{(k)} \tag{64}$$

as these diagonal elements. Then, evaluating the trace at this variation gives

$$\mathrm{tr}\left(\Theta^{(k)}\overline{\delta\Sigma}^{(k)}\right) \geq 0. \tag{65}$$

This shows that the variation $\overline{\delta\Sigma}^{(k)}$ is in an improving direction. Furthermore, we are guaranteed that the incomplete data log likelihood is nondecreasing as a result of the M-step of the expectation-maximization algorithm because, at this step, the conditional expectation of the complete data log likelihood, given the incomplete data and the last iterate for Σ_p, is maximized over Σ_p. As shown in [20] and [21], this implies that the incomplete data log likelihood is nondecreasing.

Lemma 3: Assume that $N_0 > 0$ and $\Sigma_p^{(0)}$ is positive definite. Then 1) $\Sigma_p^{(k)}$ is positive definite for all k; and 2) all stable points satisfy the trace condition (27) for all admissible variations (60).

The proof of the first part of Lemma 3 is in the Appendix. For the second part, since the diagonal elements of $\Sigma_p^{(k)}$ are positive, (65) holds with equality if and only if the diagonal elements of $\Theta^{(k)}$ are zero. Notice that if $\Sigma_p^{(k+1)} = \Sigma_p^{(k)}$, then the diagonal elements (64) are zero. This implies that the diagonal elements of $\Theta^{(k)}$ are zero and hence that

$$\mathrm{tr}\left(\Theta^{(k)}\delta\Sigma\right) = 0 \tag{66}$$

for all diagonal $\delta\Sigma$. Thus all stable points satisfy the trace condition (27) for admissible variations. From (55), stable points of the sequence $\Sigma_p^{(k)}$ yield stable points of the sequence of conditional-mean estimates of the reflectance process.

Computational Considerations

The computations required to produce radar images with our method are specified by (54)–(56). The number of iterations of these equations required to produce an image near the convergence point is presently unknown. Our experience in using an iterative algorithm to produce maximum likelihood images for emission tomography suggests that 50–100 iterations may be necessary, but this is only a guess that will not be verified until some experiments are completed. Some form of specialized processor to accomplish each iteration stage efficiently will probably be needed to produce images in practically useful times. One possible approach is the following. The matrix product

$$\Gamma = \sqrt{2E_T}M_{CR}W_P M_R S$$

is required at each iteration stage and does not change. This $I_R I_{CR} \times N$-dimensional matrix can therefore be computed once off-line, stored, and then used as needed during on-line computations. Then, at iteration stage k, the following on-line computations can be performed:

1) compute the $N \times N$-dimensional matrix A defined by $A = \Gamma^+\Sigma_p^{(k)}\Gamma + N_0 I$;
2) compute the $I_R I_{CR} \times N$-dimensional matrix B defined by $B = \Sigma_p^{(k)}\Gamma$;
3) compute $BA^{-1}r$ and the diagonal elements of $\Sigma_p^{(k)} - BA^{-1}B^+$.

The computations in 3) can be accomplished in about $4N + I_R I_{CR} - 2$ time steps using the systolic array described by Comon and Robert [26] augmented, as they suggest, by one row to accomplish the postmultiplication of BA^{-1} by r and by $I_R I_{CR}$ rows to accomplish the postmultiplication by B^+. The matrix multiplications in 1 and 2 for determining A and B can also be performed rapidly on a systolic array. More study of implementation approaches is needed, but it does not appear that the computational complexity of our new imaging algorithm needs to be a limitation to its practical use.

The choice of N, I_R and I_{CR} is important for the computations. These parameters are selected to achieve a desired range and cross-range resolution and are, therefore, problem dependent, but the same considerations used with other approaches to radar imaging can be used in selecting them. Choosing the product $I_R I_{CR}$ to be of the order of N will, in some sense, make the imaging problem well-defined because the number of unknown parameters $I_R I_{CR}$ that need to be estimated to form the image is then comparable to the number of measurements N. On the other hand, the choice of P is unique to our approach. As stated, we need $P \geq N$, but no upper limit is given. In [24], it is shown that as P increases toward infinity, so does the maximum value of the incomplete data log likelihood function, with probability one. Thus P cannot be made arbitrarily large from a theoretical standpoint. Any computation involving a matrix with one dimension equal to P can be performed off-line.

IV. Conclusion

The expressions we have obtained for forming images of diffuse, fluctuating radar targets are based on the model stated in Section II. The target reflectance is assumed to introduce wide-sense-stationary uncorrelated scattering of the transmitted signal with no glint or specular components present. The reflectance process is assumed to be a

WSSUS Gaussian process with unknown second-order statistics given by a delay-dependent covariance or scattering function. Echoes of the transmitted signal are received from all the reflecting patches that make up the target, with each patch introducing some propagation delay, Doppler shift, and random amplitude-scaling into the signal it reflects. The superposition of the echoes from all the patches is received in additive noise. Thus the reflectance process is only observed indirectly, following a linear superposition and in additive noise; thus neither the reflectance process nor its second-order statistics are known. Target images are made by displaying estimates of either the reflectance process or its second-order statistics (scattering function) based on processing the received signal. In Section II, we derived the trace condition which the maximum likelihood estimate of the covariance of the reflectance must satisfy, and we concluded that this condition is too complicated to solve explicitly for the estimate. This motivated the introduction in Section III of the incomplete-complete data model and the use of the expectation-maximization algorithm, which results in a sequence of estimates of the scattering function having increasing likelihood. A corresponding sequence of estimates of the reflectance process is also obtained.

A number of issues have yet to be resolved for the approach to radar imaging we have presented. One of the most important is resolving how glint and specular components in the return echoes should be modeled and accommodated in the formation of the images. The selection of transmitted signals to produce good images is an important subject about which little study has been made. The quality of target images obtained with our new approach is not known at present; to study this issue, we are presently implementing a computer simulation so that comparisons to alternative processing strategies can be made. The equations we have developed are computationally demanding because the matrices involved can be of large dimension and the iteration must be performed repeatedly until a stable point is approached. It is therefore important to determine the conditions under which our approach yields radar images of sufficiently improved quality compared to existing approaches to warrant the development of special processing architectures that will make it practical. The computer simulations should be of some help in this. At this time, however, the only experimental results suggesting the efficacy of our method are those reported in [21] and [22] for estimating power-density spectra in one dimension.

Acknowledgment

We are grateful to Dr. Richard E. Blahut of IBM, Owego, NY, for reading a draft of this paper and making several helpful suggestions for improving it and to Dr. Jeffrey Shapiro of MIT, Cambridge, MA, for bringing the work of Gaarder [9] to our attention.

Appendix

Proof of Lemma 1

From the definition of the log likelihood function in (26), we have

$$\frac{1}{\alpha}\left(L_{id}(K+\alpha\delta K)-L_{id}(K)\right) = -\frac{1}{\alpha}r^{+}\left(\left(K_r+\alpha 2E_TS^{+}\delta KS\right)^{-1}-K_r^{-1}\right)r - \frac{1}{\alpha}\ln\left(\det\left(K_r+\alpha 2E_TS^{+}\delta KS\right)\det\left(K_r^{-1}\right)\right) \quad \text{(A1)}$$

where K_r is the covariance of the incomplete data r as given in (26). Examining the first term on the right, we have

$$-\frac{1}{\alpha}r^{+}K_r^{-1}\left(\left(I+\alpha 2E_TS^{+}\delta KSK_r^{-1}\right)^{-1}-I\right)r = \frac{1}{\alpha}r^{+}K_r^{-1}\left(\alpha 2E_TS^{+}\delta KSK_r^{-1}\right)\left(I+\alpha 2E_TS^{+}\delta KSK_r^{-1}\right)^{-1}r = r^{+}K_r^{-1}2E_TS^{+}\delta KSK_r^{-1}r+O(\alpha). \quad \text{(A2)}$$

Examining the second term on the right in (A1), we have

$$-\frac{1}{\alpha}\ln\left(\det\left(I+\alpha 2E_TS^{+}\delta KSK_r^{-1}\right)\right) = -\frac{1}{\alpha}\ln\left(\det\left(I+\alpha B\right)\right) = -\frac{1}{\alpha}\ln\left(1+\alpha\,\mathrm{tr}(B)+\cdots+\alpha^{n}\det(B)\right) = -\mathrm{tr}(B)+O(\alpha) \quad \text{(A3)}$$

where B is defined in the first equality. For any $K\in\Omega$ to be a local maximum, a small variation in K in an admissible direction cannot increase $L_{id}(K)$, or

$$\lim_{\alpha\downarrow 0^{+}}\frac{1}{\alpha}\left(L_{id}(K+\alpha\delta K)-L_{id}(K)\right)\le 0 \quad \text{(A4)}$$

for all admissible variations δK. If K is an interior point, then $-\delta K$ is also an admissible variation and (A4) becomes an equality. Substituting (A2) and (A3), we get

$$r^{+}K_r^{-1}2E_TS^{+}\delta KSK_r^{-1}r-\mathrm{tr}\left(2E_TS^{+}\delta KSK_r^{-1}\right)=0, \quad \text{(A5)}$$

which is the trace condition (27).

Proof of Lemma 2

Suppose that K satisfies (27) and (29) for all admissible variations δK. We now show that (29) is simply the second derivative of $L_{id}(K)$ in the direction δK by taking the limit

$$\lim_{\alpha\to 0}\frac{1}{|\alpha|}\mathrm{tr}\Big(\left(2E_TS^{+}(K+\alpha\delta K)S+N_0I\right)^{-1} \cdot\left(rr^{+}-N_0I-2E_TS^{+}(K+\alpha\delta K)S\right) \cdot\left(2E_TS^{+}KS+\alpha 2E_TS^{+}\delta KS+N_0I\right)^{-1}S^{+}\delta KS - K_r^{-1}\left(rr^{+}-N_0I-2E_TS^{+}KS\right)K_r^{-1}S^{+}\delta KS\Big) = \mathrm{tr}\Big(K_r^{-1}2E_TS^{+}\delta KSK_r^{-1}\left(2E_TS^{+}KS + N_0I-2rr^{+}\right)K_r^{-1}2E_TS^{+}\delta KS\Big). \quad \text{(A6)}$$

Thus the conditions of Lemma 2 are the standard sufficient conditions for a point K to be the local maximum of $L_{id}(K)$. Equation (27) says that the first derivative of $L_{id}(K)$ is zero at K. Equation (29) says that the second derivative is negative definite along admissible variations from K. A necessary condition for K to be a relative maximum is that this last expression evaluated at K to be equal to or less than zero for all admissible variations δK. Under the assumptions in Section IV, admissible variations are given by (60). Substituting (60) into (A6) and evaluating for all diagonal matrices $\delta\Sigma$ gives the second-order necessary condition.

Estimating a General Toeplitz Matrix

In Section IV, we derived a sequence of estimates for a covariance matrix subject to the constraint that the estimates be circulant Toeplitz matrices. For completeness, we develop and discuss the equations for estimating a covariance matrix subject to the weaker constraint that the estimates be general Toeplitz matrices. Similar equations for other constraints on the Toeplitz matrices are easily obtained by mimicking the steps in the main body of this paper.

Let the complete data be $\{b, w\}$, and let b be normally distributed with zero mean and covariance K, as given in (27). The complete data log likelihood is

$$L_{cd}(K_p) = -\ln(\det(K_p)) - b_p^+ K_p^{-1} b_p \tag{A7}$$

where all terms that are not a function of K_p have been suppressed. Maximizing this function gives the trace condition

$$\mathrm{tr}\left(K^{-1}(bb^+ - K)K^{-1}\delta K\right) = 0, \tag{A8}$$

which the maximum likelihood estimate K must satisfy. Performing the E- and M-steps of the expectation-maximization algorithm yields the following iteration sequence for the elements $K(n,i)$, $n = 0,1,\cdots,N-1$, of the covariance matrix $K(i)$ defined in (23):

$$K^{(k+1)}(n,i) = \frac{1}{N-n}\mathrm{E}\left[\sum_{j=0}^{N-n-1} b(j,m+i)\,b^*(j+n,m+i)\,|r, K^{(k)}\right]. \tag{A9}$$

In matrix form,

$$K^{(k+1)} = K^{(k)} + 2E_T K^{(k)} S K_r^{(k)-1} \cdot\left(rr^+ - 2E_T S^+ K^{(k)} S - N_0 I\right) K_r^{(k)-1} S^+ K^{(k)} \tag{A10}$$

where

$$K_r^{(k)} = 2E_T S^+ K^{(k)} S + N_0 I. \tag{A11}$$

If this iteration converges to a stable point, then the trace condition is satisfied at this point, as may be shown by using the same arguments as in Section IV. It is worth restating that the reason this iteration is not recommended here is that the probability that the iteration sequence generates a singular estimate for K approaches one as N gets large. By restricting consideration to Toeplitz matrices with circulant extensions, the log likelihood function is bounded with probability one for finite extensions and a positive definite K is generated with probability one, as proven by Fuhrmann and Miller [24].

Proof of Lemma 3, Part 1

Assume that the initial guess $\Sigma_p^{(0)}$ for Σ_p is positive definite and that $N_0 > 0$. We will now show that, if $\Sigma_p^{(k)}$ is positive definite, then so is $\Sigma_p^{(k+1)}$, and thus, by induction, $\Sigma_p^{(k)}$ is positive definite for all k. One key to following this derivation is the matrix identity

$$B(I + AB)^{-1} = (I + BA)^{-1}B. \tag{A12}$$

This identity is used to rewrite (57) as

$$\begin{aligned}\Sigma_p^{(k+1)} &\stackrel{d}{=} H\left(N_0\Sigma_p^{(k)}\Gamma\Gamma^+\Sigma_p^{(k)} + N_0^2\Sigma_p^{(k)} + \Sigma_p^{(k)}\Gamma rr^+\Gamma^+\Sigma_p^{(k)}\right)H^+ \\ &= N_0\left(H\Sigma_p^{(k)}\Gamma\right)\left(\Gamma^+\Sigma_p^{(k)}H^+\right) + \left(H\Sigma_p^{(k)}\Gamma r\right)\left(r^+\Gamma^+\Sigma_p^{(k)}H^+\right) + N_0^2 H\Sigma_p^{(k)}H^+ \end{aligned} \tag{A13}$$

where we have defined H according to

$$H = \left(\Sigma_p^{(k)}\Gamma\Gamma^+ N_0 I\right)^{-1}. \tag{A14}$$

Clearly, all the diagonal elements of (A13) are greater than or equal to zero. To show that they are strictly positive, we look at the last term and get that the ith diagonal element is

$$\begin{aligned}\left(N_0^2 H\Sigma_p^{(k)}H^+\right)_{ii} &= N_0^2\sum_{j=0}^{I_R I_{CR}-1}(H)_{i,j}\left(\Sigma_p^{(k)}\right)_{jj}(H^+)_{ji} \\ &= N_0^2\sum_{j=0}^{I_R I_{CR}-1}|(H)_{i,j}|^2\left(\Sigma_p^{(k)}\right)_{jj},\end{aligned} \tag{A15}$$

which is clearly positive when $N_0 > 0$ since H is invertible and all diagonal elements of $\Sigma_p^{(k)}$ are positive.

REFERENCES

[1] D. R. Wehner, *High Resolution Radar*. Dedham, MA: Artech House, 1987.

[2] D. L. Mensa, *High-Resolution Radar Imaging*. Dedham, MA: Artech House, 1984.

[3] M. Bernfeld, "CHIRP Doppler radar," *Proc. IEEE*, vol. 72, pp. 540–541, Apr. 1984.

[4] D. L. Snyder, H. J. Whitehouse, J. T. Wohlschlaeger, and R. Lewis, "A new approach to radar/sonar imaging," in *Proc. SPIE Conf. Advanced Algorithms and Architectures*, vol. 696, San Diego, CA, Aug. 1986.

[5] J. L. Walker, "Range-Doppler imaging of rotating objects," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, no. 1, pp. 23–52, Jan. 1980.

[6] J. Shapiro, B. A. Capron, and R. C. Harney, "Imaging and target detection with a heterodyne-reception optical radar," *Appl. Opt.*, vol. 20, no. 19, pp. 3292–3313, Oct. 1981.

[7] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, vol. 3. New York: Wiley, 1971.

[8] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Machine Intel.*, vol. PAMI-4, no. 2, pp. 643–652, Mar. 1982.

[9] N. T. Gaarder, "Scattering function estimation," *IEEE Trans. Inform. Theory*, vol. IT-14, no. 5, pp. 684–693, Sept. 1968.

[10] P. E. Green, "Radar measurements of target scattering properties," in *Radar Astronomy*, V. V. Evans and T. Hagfors, Eds. New York: McGraw-Hill, 1968, ch. 1, pp. 1–78.

[11] T. Kailath, "Measurements in time-variant communication channels," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 829–836, Sept. 1962.

[12] R. G. Gallager, "Characterization and measurement of time-and-frequency spread channels," MIT Lincoln Laboratory, Lexington, MA, Tech. Rep. 352, Apr. 1964.

[13] T. Hagfors, "Some properties of radio waves reflected from the moon and their relation to the lunar surface," *J. Geophys. Res.*, vol. 66, p. 777, 1961.

[14] ——, "Measurement of properties of spread channels by the two-frequency method with application to radar astronomy," MIT Lincoln Laboratory, Lexington, MA, Tech. Rep. 372, Jan. 1965.

[15] R. Price, "Maximum-likelihood estimation of the correlation function of a threshold signal and its application to the measurement of the target scattering function in radar astronomy," MIT Lincoln Laboratory Group, Lexington, MA, Group Rep. 34-G-J, May 1962.

[16] M. V. Levin, "Estimation of the second-order statistics of randomly time-varying linear systems," MIT Lincoln Laboratory, Lexington, MA, Group Rep. 34-G-7, Nov. 1962.

[17] L. G. Abraham *et al.*, "Tropospheric-scatter propagation tests using a RAKE receiver," in *1965 IEEE Communications Conv. Rec.*, 1965, pp. 685–690.

[18] B. M. Sifford *et al.*, "HF time- and frequency-dispersion effects: Experimental validation of an FSK error rate model," Stanford Res. Inst., Menlo Park, CA, Tech. Rep. 4, Mar. 1965.

[19] B. Reiffen, "On the measurement of atmospheric multipath and Doppler spread by passive means," MIT Lincoln Laboratory, Lexington, MA, Tech. Note 1965-6 G-66, Mar. 1965.

[20] A. D. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc.*, vol. B, 39, pp. 1–37, 1977.

[21] M. I. Miller and D. L. Snyder, "The role of likelihood and entropy in incomplete-data problems: Applications to estimating point-process intensities and Toeplitz-constrained covariances," *Proc. IEEE*, vol. 75, no. 7, July 1987.

[22] M. Turmon and M. I. Miller, "Performance evaluation of maximum-likelihood estimates of Toeplitz covariances generated with the expectation-maximization algorithm," in *Proc. 1987 Conf. Information Sciences and Systems*, Johns Hopkins Univ., Baltimore, MD, pp. 25–27.

[23] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proc. IEEE*, vol. 70, no. 9, pp. 963–974, Sept. 1982.

[24] D. R. Fuhrmann and M. I. Miller, "On the existence of positive definite maximum-likelihood estimates of structured covariance matrices," *IEEE Trans. Inform. Theory*, to appear.

[25] S. A. Tretter, *Introduction to Discrete-Time Signal Processing*. New York: Wiley, 1976.

[26] P. Comon and Y. Robert, "A systolic array for computing BA^{-1}," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-35, no. 6, pp. 717–723, June 1987.

6.2 Reprint of: P. Moulin, D. Snyder, and J. O'Sullivan, "Maximum-Likelihood Spectrum Estimation of Periodic Processes from Noisy Data," Proc. 1989 CISS Conference, Johns Hopkins University, Baltimore, MD, March 1989.

Maximum-Likelihood Spectrum Estimation of Periodic Processes from Noisy Data *

P. Moulin
D. L. Snyder
J. A. O'Sullivan

Electronic Systems and Signals Research Laboratory
Department of Electrical Engineering
Washington University
Saint Louis, MO 63130

ABSTRACT

We have developed a new approach to maximum-likelihood spectrum estimation of wide-sense stationary processes from noisy data. A statistical model for the data is defined. The process whose spectrum is sought is wide-sense stationary, periodic and Gaussian, and its observations are corrupted by an additive white noise. A maximum-likelihood formulation of this problem has been derived, and the equations are solved numerically via the expectation-maximization algorithm. This approach presents several attractive features, an important one being that the noise corrupting the observations is now taken into account.

We present some recent developments for this problem. The statistical performance of the new maximum-likelihood spectrum estimator is studied both theoretically and numerically. Comparison with traditional estimators such as the periodogram highlight several strong points of the method. We also identify certain limitations, namely the instability of estimates for high noise levels. These limitations can be alleviated if *a priori* information about the signal is available. Two such problems are discussed in which the information at hand has the form of a constraint on the input signal-to-noise ratio.

We show how such information can be incorporated in the maximum-likelihood estimation procedure. First we assume the signal power to be known. Theoretical issues of existence and uniqueness of the solution are discussed. We proceed with a problem in which the information is less complete, when only an upper bound and/or a lower bound on the signal power are available. The statistical performance of both constrained estimators is quantitatively studied.

* This work was supported by contract number N00014-86-K-0370 from the Office of Naval Research.

Maximum-Likelihood Spectrum Estimation of Periodic Processes from Noisy Data *

P. Moulin
D. L. Snyder
J. A. O'Sullivan

Electronic Systems and Signals Research Laboratory
Department of Electrical Engineering
Washington University
Saint Louis, MO 63130

1. Introduction

A promising approach to maximum-likelihood estimation of Toeplitz constrained covariance matrices has been proposed recently [1]. Several further developments can be considered. First, this method also applies to the dual problem of spectrum estimation. Another issue of interest is that the statistical model can account for the presence of additive noise corrupting the observations and for linear transformations of the process whose covariance or spectrum is sought. These considerations have motivated a new approach to high-resolution delay-doppler radar imaging, where a major goal is to produce estimates of the target's scattering function [2]. In the special case of a point target and a constant envelope transmitted signal, this reduces to a spectrum estimation problem.

This paper describes some recent developments for this problem. We study the *statistical performance* of the new maximum-likelihood spectrum estimator both theoretically and numerically. Comparison with traditional estimators such as the periodogram highlight several strong points of the method. We also identify certain limitations, namely the instability of estimates for high noise levels. These limitations can be alleviated if *a priori* information about the signal is available. Two such problems are discussed here in which the information at hand has the form of a constraint on the input signal-to-noise ratio.

This paper is organized as follows. Our model is presented in Section 2. A maximum-likelihood formulation of the problem is given in Section 3, and the equations are solved via the expectation-maximization algorithm. Section 4 is devoted to a statistical performance analysis of this estimator and a comparison with two other methods. In Section 5 we show how *a priori* information on the signal can be incorporated in the maximum-likelihood estimation procedure. First we assume the signal power to be known. Theoretical issues of existence and uniqueness of the solution are discussed. Section 5 deals with a less complete knowledge, where only an upper bound and/or an lower bound on the signal power are available. The last section is devoted to a quantitative study of the statistical performance of both constrained estimators.

2. Model

The following is derived from the model presented in [1] for a point target and a constant envelope transmitted signal. The observation is an N-vector sample of a wide-sense stationary, periodic, Gaussian process corrupted by an additive noise :

$$r = b + w \ , \tag{2.1}$$

where b contains N consecutive samples of a zero-mean periodic process b_p with length $P \geq N$, and w is an zero-mean white Gaussian noise with variance N_0, uncorrelated with b. The periodicity assumption is required to guarantee that the likelihood function is bounded above; therefore, there exists the maximum-likelihood estimator [1].

* This work was supported by contract number N00014-86-K-0370 from the Office of Naval Research.

Now we define the spectral process associated with b_p to be the DFT of one period of b_p. Assume that we are interested in estimating only M of the components of this spectral P-vector ($1 \le M \le P$), the other components being zero with probability 1; let c be this M-vector. This assumption is introduced to deal with the bandlimited spectra encountered in radar applications, which arise because radar targets have finite extent [2]. c is a Gaussian random M-vector with diagonal covariance Σ, whose entries $\sigma^2(i)$, i = 0,..,M-1, are real and positive. c and b are related by a linear transformation :

$$b = \Gamma^\dagger c \ , \tag{2.2}$$

where we have defined the $M \times N$ matrix Γ, consisting of the first N rows and the outer M columns of the $P \times P$ DFT matrix. The superscript † denotes the Hermitian-transpose operator on matrices. Our model for the observations can now be written as

$$r = \Gamma^\dagger c + w \ . \tag{2.3}$$

The covariance matrix for r is given by

$$K_r = E[rr^\dagger] = \Gamma^\dagger \Sigma \Gamma + N_0 I_N \ , \tag{2.4}$$

where I_N is the $N \times N$ identity matrix.

3. Spectrum Estimators

In this section we introduce a maximum-likelihood spectrum estimator for the model (2.3), denoted by ML1. We also define two estimators which will be analyzed and compared to ours in the next section. The first one is the maximum-likelihood estimator derived assuming noise-free data, denoted by ML0 ; the second one is the periodogram.

3.1. ML1 Estimator

From (2.4), the likelihood function for Σ is

$$L(r,\Sigma) = -\tfrac{1}{2} \ln \det (\Gamma^\dagger \Sigma \Gamma + N_0 I_N) - \tfrac{1}{2}\, r^\dagger (\Gamma^\dagger \Sigma \Gamma + N_0 I_N)^{-1} r \ . \tag{3.1}$$

Maximizing the likelihood with respect to Σ yields the necessary trace condition which the estimate $\hat{\Sigma}$ must satisfy [1,2]:

$$Tr\ [\Gamma (\Gamma^\dagger \hat{\Sigma} \Gamma + N_0 I_N)^{-1} (rr^\dagger - \Gamma^\dagger \hat{\Sigma} \Gamma - N_0 I_N)(\Gamma^\dagger \hat{\Sigma} \Gamma + N_0 I_N)^{-1} \Gamma^\dagger \delta\hat{\Sigma}] = 0 \ , \tag{3.2}$$

for all $M \times M$ diagonal matrices $\delta\hat{\Sigma}$. This trace condition is a nonlinear equation in $\hat{\Sigma}$. Generally it cannot be solved directly in closed-form, so some numerical search procedure must be implemented. An elegant solution is the expectation-maximization (EM) algorithm used in [1,2]. An initial estimate $\hat{\Sigma}^{(0)}$ is selected. At step k+1 (k = 0,1,..) the estimate is updated according to

$$\hat{\Sigma}^{(k+1)} = argmax\ Q(\Sigma | \hat{\Sigma}^{(k)}) \tag{3.3}$$

where

$$Q(\Sigma | \hat{\Sigma}^{(k)}) = -\tfrac{1}{2} \sum_{i=0}^{M-1} \ln \sigma^2(i) - \tfrac{1}{2} \sum_{i=0}^{M-1} \frac{E[|c(i)|^2 | r, \hat{\Sigma}^{(k)}]}{\sigma^2(i)} \ , \tag{3.4}$$

and

$$E[|c(i)|^2 | r, \hat{\Sigma}^{(k)}] = [\hat{\Sigma}^{(k)} - \hat{\Sigma}^{(k)} \Gamma (\Gamma^\dagger \hat{\Sigma}^{(k)} \Gamma + N_0 I_N)^{-1} \Gamma^\dagger \hat{\Sigma}^{(k)} + \hat{\Sigma}^{(k)} \Gamma (\Gamma^\dagger \hat{\Sigma}^{(k)} \Gamma + N_0 I_N)^{-1}$$
$$\times rr^\dagger (\Gamma^\dagger \hat{\Sigma}^{(k)} \Gamma + N_0 I_N)^{-1} \Gamma^\dagger \hat{\Sigma}^{(k)}]_{ii} \ . \tag{3.5}$$

This algorithm produces a sequence of estimates

$$\hat{\sigma}^2(i)^{(k+1)} = E[|c(i)|^2 | r, \hat{\Sigma}^{(k)}] \tag{3.6}$$

having increasing likelihood. It can be shown that the stable points of this algorithm satisfy the necessary trace condition for a maximizer [2]. The issue of uniqueness is addressed in [3].

Special case : N = M = P

In this special case, a full period of the process is observed. Γ is equal to the $P \times P$ DFT matrix W_p. A closed-form expression for $\hat{\Sigma}$ can be derived:

$$\hat{\sigma}^2(i) = \max(0, |(W_p r)(i)|^2 - N_0) . \tag{3.7}$$

3.2. ML0 Estimator

Additive noise corrupting observations is usually not included separately in approaches to spectrum estimation. This model was assumed in [1]. The sequence of estimates of Σ is still given by (3.6) and (3.5), in which we now let $N_0 = 0$. We call this the ML0 estimation. Clearly ML0 and ML1 are equivalent for noise-free problems.

Special case : N = M

The problem for which the number of observations (N) is equal to the number of parameters to be estimated (M) is of some practical interest. It also turns out that the trace condition can be solved in closed-form in this instance. The matrix Γ is then invertible, indicating the existence of a one-to-one mapping between r and c. The ML0 estimator is simply

$$\hat{\sigma}^2(i) = |(\Gamma^{-t} r)(i)|^2 , \tag{3.8}$$

where Γ^{-t} denotes $(\Gamma^{-1})^t$.

3.3. Periodogram

The periodogram estimate of the spectrum is defined as the (scaled) magnitude-squared Fourier transform of the N observations padded with P-N zeroes [3]. The first M spectrum samples are then given by

$$\hat{\sigma}^2(i) = (P/N) |(\Gamma r)(i)|^2 \tag{3.9}$$

Special case : N = M = P

When N = M = P, the periodogram and ML0 estimates are the same.

4. Bias Performance Analysis

4.1. Performance Evaluation

In this section, we estimate Σ for the model (2.3) and study the statistical performance of the three estimators above. For each method the bias is evaluated, where

$$Bias[\hat{\Sigma}] = E[\hat{\Sigma}] - \Sigma \tag{4.1}$$

As we shall see in Section 4.3, the performance strongly depends upon the input signal to noise ratio defined by

$$SNR_{in} = E_0 / N_0 , \tag{4.2}$$

where E_0 is the average power of the process, defined by

$$E_0 = (1/P)\, Tr\, [\Sigma] . \tag{4.3}$$

In sections 4.2 and 4.3, we evaluate the bias for the estimators derived in Section 3. Whenever closed-form expressions are not available, computer simulations are performed. Typically 3000 realizations are generated for each process. For a given estimator, (4.1) is then estimated from the 3000 estimates. The analysis is carried out at various input SNR levels. Much effort was made for the special case M = N. This provides insight into the problem since the ML0 equations can be solved in closed form. The choice of P is free, so long as $P \geq N$ [2].

4.2. Closed-form Expressions for Estimator Bias Performance

(a) ML1

As indicated in Section 3.1, no closed-form expression for the estimator is available, so the evaluation of the bias is obtained by computer simulation.

(b) ML0

Closed-form expressions for ML0 can be derived when M = N. The results are presented below. Combining (2.3) and (3.8), we can write

$$\hat{\sigma}^2(i) = |(c + \Gamma^{-t} w)(i)|^2 . \tag{4.4}$$

Taking the expectation of (4.4), we get

$$E[\hat{\sigma}^2(i)] = \sigma^2(i) + N_0 (\Gamma\Gamma^t)^{-1}_{ii} , \tag{4.5}$$

which implies

$$Bias[\hat{\sigma}^2(i)] = N_0 (\Gamma\Gamma^t)^{-1}_{ii} . \tag{4.6}$$

The bias is due to the noise corrupting the observations and is proportional to its variance. The sensitivity of the bias to the noise is determined by the diagonal entries of the matrix $(\Gamma\Gamma^t)^{-1}$.

(c) Periodogram

Bias

Combining (2.3) and (3.9), we write the periodogram estimates in the equivalent form

$$\hat{\sigma}^2(i) = (P/N) |(\Gamma\Gamma^t c + \Gamma w)(i)|^2 . \tag{4.7}$$

Taking the expectation of (4.7), we get

$$E[\hat{\sigma}^2(i)] = (P/N) (\Gamma\Gamma^t \Sigma \Gamma\Gamma^t)_{ii} + N_0 , \tag{4.8}$$

and

$$Bias[\hat{\sigma}^2(i)] = [(P/N) (\Gamma\Gamma^t \Sigma \Gamma\Gamma^t)_{ii} - \sigma^2(i)] + N_0 . \tag{4.9}$$

The bias contains two terms. The second is due to the noise and is proportional to N_0. The other term is independent of N_0. Even for noise-free observations, the periodogram is a biased estimator of Σ unless $\Gamma\Gamma^t$ is the identity matrix. This would be the case only for N = M = P (observation of a full period of the process) or N/M $\rightarrow \infty$ (infinite data).

4.3. Simulation results

Process 1

The first process we consider is real and has period P = 10. Its spectrum is symmetric and lowpass (M = 5). All nonzero spectrum samples are identical :

$$\sigma^2(i) = 1 , \qquad i = 0,..,4 .$$

The number of observations is N = M = 5. The noise variance N_0 ranges from 0 to 1. Figure 1 shows the bias for the estimators of $\sigma^2(2)$ as a function of SNR_{in}, according to the definitions (4.1) and (4.2). In the absence of additive noise ($SNR_{in} \rightarrow \infty$), ML1 and ML0 are the same. Both are unbiased estimators. The periodogram, however, is biased. When N_0 increases from 0, the performance of the estimators is roughly constant so long as SNR_{in} remains above some threshold. For larger N_0, all three estimators exhibit a strong degradation in performance. Comparing the thresholds for ML0 and ML1, we see the tremendous improvements brought by taking the noise into account in the model. Typically, for a same bias, ML1 will have the same performance as ML0 operating in a 20 dB noisier environment.

Process 2

As shown in the next section, the periodogram does not perform well for rough spectra. This motivated our study of a sharply peaked spectrum. The process has period P = 10, and a single nonzero spectrum component

$$\sigma^2(0) = 1 .$$

There is just $N = M = 1$ observation.

The bias for the estimators of $\sigma^2(0)$ is plotted as a function of SNR_{in} in Figure 2. In the absence of additive noise, the bias of the periodogram is -90% of $\sigma^2(0)$. Clearly, the conventional estimator is outperformed by ML0 and ML1. It should also be noticed that for this process, the improvement of ML1 over ML0 is quite reduced.

Computational Considerations

The convergence rate of the EM algorithm depends on several parameters. The computation time for each iteration is of order $M\ N^2$. The number of iterations required for convergence of the algorithm grows as M and N increase. For ML1, more iterations are required as N_0 increases, especially in the threshold region and beyond. Typical figures are: for process 1 with $N_0 = 0.1$, 30 iterations are required before the spectrum estimates are stable; when $N_0 = 1$, 300 iterations must be performed. Our algorithm is implemented on a Masscomp model 5500. Running the program on 3000 realizations in the latter case is typically completed in 6 CPU hours. We are presently implementing these algorithms on a mesh-connected 1024 processor (DAP by Active Memory Technology), and we expect a major reduction in the time required to produce estimates.

4.4. Discussion

The results derived above suggest additional comments on the periodogram. It can be shown that (4.8) can be written in the alternative form [4]

$$E[\hat{\sigma}^2(i)] = \sigma^2(i) * \Omega_N(i) + N_0 , \tag{4.10}$$

where * denotes the discrete convolution operation, and $\Omega_N(i)$ is the DFT of the window

$$\begin{aligned} \omega_N(n) &= 1 - \frac{|n|}{N} \quad : |n| < N \qquad \text{for } P \geq 2N \\ &= 0 \qquad\quad : N \leq |n| < \frac{P}{2} \\ \omega_N(n) &= 1 - \frac{|n|}{N} \quad : |n| < P - N \qquad \text{for } 2N > P \geq N \\ &= \frac{P}{N} \qquad : P - N \leq |n| < \frac{P}{2} \end{aligned} \tag{4.11}$$

The main lobe of Ω_N has width $\frac{P}{N}$. Consider now a process made of sharp isolated spectral peaks, such as Process 2 above. Equation (4.10) shows that the energy in these peaks is spread out as a result of the convolution operation ; consequently, these peaks are grossly estimated when $\frac{P}{N}$ is large.

5. Constrained maximum-likelihood estimation

5.1. Description of the problem

In this section we focus our attention on ML1. An examination of Figures 1-4 suggests that ML1 suffers in certain situations. When SNR_{in} is low, the estimates are biased ; as we shall soon see, their variance is also large. Although the maximum-likelihood estimator is asymptotically unbiased and efficient, these properties are not guaranteed in the small-sample problems considered in Section 4. This limitation can be alleviated if *a priori* knowledge, such as SNR_{in}, is available. Since N_0 is known, such a constraint on the signal-to-noise ratio can be translated into a constraint on the signal power that must be satisfied by

the maximum-likelihood estimates. Now we show how this constraint can be incorporated into the EM algorithm. The constrained estimates exist and are unique.

In Section 5.2, SNR_{in} is known. In Section 5.3, our knowledge is more incomplete, and only an upper bound and/or a lower bound on SNR_{in} are available.

5.2. Known SNR_{in}

The equations for ML1 presented in Section 3.1 can be modified as follows to satisfy the constraint. At each step of the EM algorithm, we maximize $Q(\Sigma|\hat{\Sigma}^{(k)})$ defined in (3.4), subject to the power constraint

$$\sum_{i=0}^{M-1} \hat{\sigma}^2(i) = PE_0 = S \ , \tag{5.1}$$

where E_0 is the signal power. The solution also maximizes

$$Q(\Sigma|\hat{\Sigma}^{(k)}) + \lambda \left(\sum_{i=0}^{M-1} \sigma^2(i) - S \right), \tag{5.2}$$

where λ is a Lagrange multiplier. Taking the gradient of (5.2) with respect to Σ, we obtain a quadratic equation for each spectral component

$$2\lambda\sigma^4(i) - \sigma^2(i) + C_i = 0 \ , \tag{5.3}$$

where

$$C_i = E[|c(i)|^2 | r, \hat{\Sigma}^{(k)}]$$

is calculated according to (3.5). The solution to (5.3) is

$$\begin{aligned} \sigma^2(i) &= \frac{1 + I_i\sqrt{1-8C_i\lambda}}{4\lambda} \quad &: \lambda \neq 0 \\ &= C_i &: \lambda = 0 \ , \end{aligned} \tag{5.4}$$

where I_i is either +1 or -1. The equation for λ is

$$4S\lambda - M = \sum_{i=0}^{M-1} I_i\sqrt{1-8C_i\lambda} \ . \tag{5.5}$$

In general this nonlinear equation in λ cannot be solved in closed-form. Furthermore, an ambiguity subsists about the choice of the signs I_i. The latter problem is solved by application of the following theorem :

Theorem

Assume that $C_0 > C_i$, i = 1,..,M-1. Then

(1)

$$\begin{aligned} I_i &= -1 \quad : i = 1,..,M-1 \\ I_0 &= +1 \quad : S < 2C_0 \left[M - \sum_{i=1}^{M-1} I_i\sqrt{1-(C_i/C_0)} \right] \\ &= -1 \quad : else \end{aligned}$$

(2) λ is the largest nonzero solution of

$$\left(4S\lambda - M + \sum_{i=1}^{M-1} I_i\sqrt{1-(C_i/C_0)} \right)^2 = 1 - 8C_0\lambda, \text{ for } S \neq \sum_{i=1}^{M-1} C_i \ , \tag{5.6a}$$

and

$$\lambda = 0, \text{ for } S = \sum_{i=1}^{M-1} C_i \ . \tag{5.6b}$$

λ is upper-bounded by $1/8C_0$, and (5.6a) can be solved numerically for λ. Note that the particular case (5.6b) is also the solution to the unconstrained maximization problem. Next, $\hat{\sigma}^2(i)^{(k+1)}$ is calculated from (5.4). The whole procedure is repeated at each maximization step of the EM algorithm. Note that because of the highly nonlinear nature of the problem, no analytic expression is available for the constrained estimator, even in the special case mentioned in (3.7).

5.3. Known upper/lower bound on SNR_{in}

In this section, the *a priori* knowledge about SNR_{in} has the form of an upper bound. Our approach parallels that of the previous section, with the upper bound now expressed as an inequality constraint on the estimated signal power. At each step of the EM algorithm, we maximize $Q(\Sigma|\hat{\Sigma}^{(k)})$ defined in (3.4), subject to the inequality constraint

$$\sum_{i=0}^{M-1} \hat{\sigma}^2(i) \le PE_{\max} = S_{\max} , \tag{5.7}$$

where $E_{\max}$ is the upper bound on the signal power. If the unconstrained solution satisfies the upper bound, the constraint is inactive and the estimate is given by (3.6). Otherwise, the constraint is active, and as in Section 5.2, the solution is the maximizer of the expression (5.2).

We can expect the performance of this estimator to be strongly conditioned by the choice of $E_{\max}$. In the limiting case $E_{\max} \to \infty$, the constraint is always inactive and the estimator is equivalent to the unconstrained estimator. For the other extreme case $E_{\max} \to 0$, the constraint is always active. A lower bound or simultaneous upper and lower bounds are treated in exactly the same manner.

6. Simulation results

In this section, we apply the SNR-constrained estimators derived above to Process 1, and we evaluate numerically both their bias and mean-squared error, where

$$Var[\hat{\Sigma}] = E[\hat{\Sigma}^2] - (E[\hat{\Sigma}])^2 \tag{6.1}$$

$$MSE\ [\hat{\Sigma}] = E[(\hat{\Sigma}-\Sigma)^2] = Var[\hat{\Sigma}] + (Bias[\hat{\Sigma}])^2 . \tag{6.2}$$

The output signal to noise ratio matrix is defined as follows :

$$SNR_{out}[\hat{\Sigma}] = E[\hat{\Sigma}]\,(MSE\ [\hat{\Sigma}])^{-1/2} . \tag{6.3}$$

Figures 3 and 4 give a plot of the bias and SNR_{out} for three different estimators of $\sigma^2(2)$ as a function of SNR_{in}, according to the definitions (4.1), (4.3), and (6.3). The estimators represented on these figures are: the two constrained estimators of Section 5, respectively denoted by EQ-MLE and INEQ-MLE, and defined for the power constraint $S = 5$ and $S_{\max} = 15$, respectively ; and the unconstrained estimator ML1 of Section 3.1.

In the absence of additive noise ($SNR_{in} \to \infty$), ML1 and EQ-MLE are unbiased. However, the periodogram is biased, and so is INEQ_MLE. For the latter, this can be understood as follows. The distribution of the sum $\hat{S}$ of the M estimates $\hat{\sigma}^2(i)$ is truncated (to $S_{\max} = 15$), and therefore the sum of all biases is negative.

When N_0 increases from 0, the performance of the estimators is roughly constant so long as SNR_{in} remains above some threshold. For larger N_0, all estimators exhibit a degradation in performance. Note that for the SNR-constrained estimators, each bias is upper-bounded by $S_c - \sigma^2(i)$, and lower-bounded by $-\sigma^2(i)$, where S_c is the constraint. Comparing the SNR_{out} performance in Figure 2, we see other favorable effects of incorporating SNR constraints into the problem. For low N_0, SNR_{out} is improved. This is due to the estimates having a lower variance, which is the dominant term in SNR_{out}. For very noisy data, the performance of the estimators is clearly improved. We can easily derive a lower bound for $SNR_{out}[\hat{\sigma}^2(i)]$:

$$\frac{\sigma^2(i)}{\max[S_c - \sigma^2(i)\,,\,\sigma^2(i)]} \le 1 .$$

This bound is independent of N_0.

Conclusions

In this paper, we have described an approach to spectrum estimation from noisy data, based upon a statistical model for the observations. First we derive a maximum-likelihood estimator, and evaluate its statistical performance. A comparison is made with two other methods that do not take the additive noise into account. One is the traditional periodogram and the other is the maximum-likelihood estimator derived for a noise-free model. It is shown that the new estimator offers a much better bias performance. The improvement over the periodogram is particularly noticeable for rough spectra: The bias of the periodogram was as high as 90% for the process #2 we considered.

In general however, the maximum-likelihood estimates are still unstable at high noise levels. In the second step of our study, we refine our technique to improve the performance when some side information exists. We have studied one such problem in which some information about the signal-to-noise ratio is available. The performance for the SNR-constrained estimators has been numerically evaluated, and compared with that of the unconstrained estimator and of the periodogram. The new estimators perform significantly better than their competitors for low SNR_{in}. Because of the SNR constraint, the estimates are not allowed to take on the large values that were produced in the unconstrained estimation problem. This results in the estimates having a lower variance. One additional feature of our approach, and an attractive one, is its versatility. Only a slight modification of the (unconstrained) algorithm is required.

References

[1] M. I. Miller, D. L. Snyder, "The Role of Likelihood and Entropy in Incomplete-Data Problems: Applications to Estimating Point-Process Intensities and Toeplitz Constrained Covariances," *Proc. IEEE, Vol.75*, No.7, July 1987.

[2] D. L. Snyder, J. A. O'Sullivan, M. I. Miller, "The Use Of Maximum-Likelihood Estimation For Forming Images Of Diffuse Radar-Targets From Delay-Doppler Data," *IEEE Trans. on Information Theory*, to appear in 1989.

[3] J. A. O'Sullivan, P. Moulin, D. L. Snyder, "Cramer-Rao Bounds for Constrained Spectrum Estimation with Application to a Problem in Radar Imaging", *Proc. of the 1988 Allerton Conference.*

[3] W. Davenport, W. Root, *Random Signals and Noise*, McGraw Hill, 1958.

[4] P. Moulin, unpublished

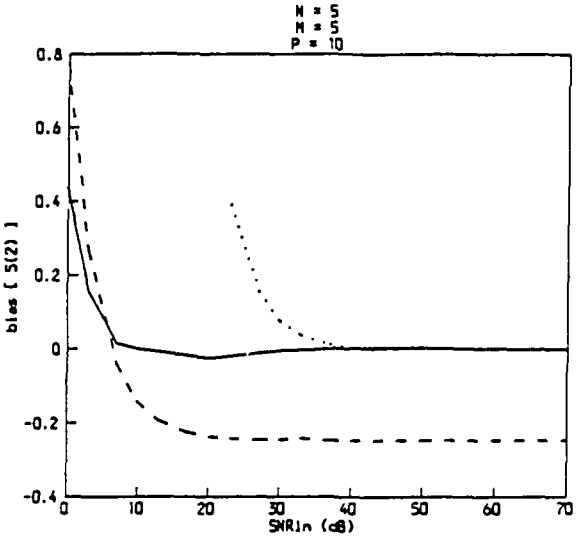


Figure 1. *Bias* ($\sigma^2(2)$) for Process 1
ML1 (solid line), ML0 (dotted line), periodogram (dashed line)

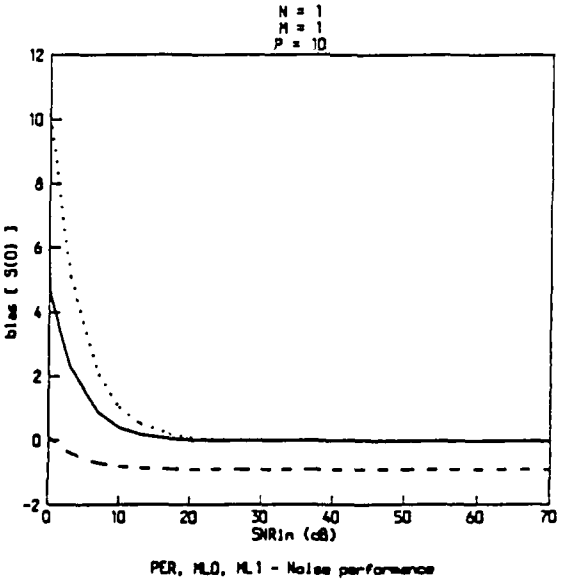


Figure 2. *Bias* ($\sigma^2(0)$) for Process 2
ML1 (solid line), ML0 (dotted line), periodogram (dashed line)

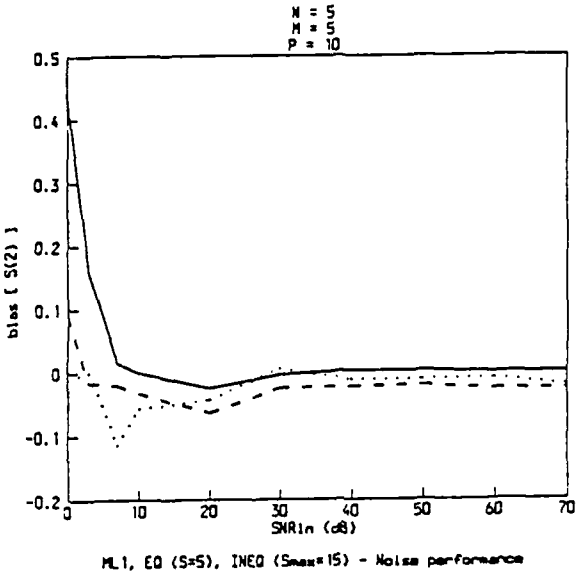


Figure 3. *Bias* ($\sigma^2(2)$)
ML1 (solid line), EQ-MLE (S=5, dotted line), INEQ-MLE (S_{max}=15, dashed line)

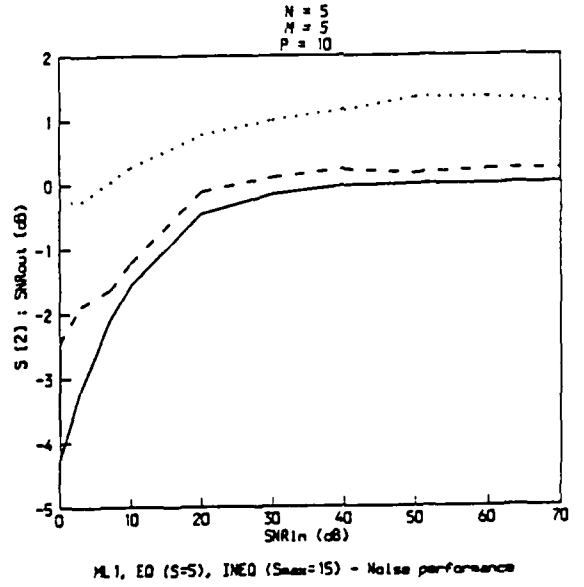


Figure 4. SNR_{out} ($\sigma^2(2)$)
ML1 (solid line), EQ-MLE (S=5, dotted line), INEQ-MLE (S_{max}=15, dashed line)

6.3 Reprint of: J. A. O'Sullivan, D. L. Snyder, and P. Moulin: "The Role of Spectrum Estimation in Forming High-Resolution Radar Images", Proc. ICASSP 1989, Glasgow, U.K., May 1989.

THE ROLE OF SPECTRUM ESTIMATION IN FORMING HIGH-RESOLUTION RADAR IMAGES

Joseph A. O'Sullivan, Donald L. Snyder, and Pierre Moulin

Electronic Systems and Signals Research Laboratory
Department of Electrical Engineering
Washington University
St. Louis, Missouri 63130

ABSTRACT

We have developed a new approach to forming high-resolution images of radar targets from delay-doppler, spotlight-mode radar data. This approach is based on a model for the target's reflectivity in terms of wide-sense stationary, uncorrelated scatterers having complex-valued Gaussian statistics. The imaging problem is to estimate the target's scattering function in terms of radar-echo data acquired with a series of target illuminations. We develop a method for solving this multidimensional spectrum estimation problem through the use of maximum-likelihood estimation implemented via the expectation-maximization algorithm.

INTRODUCTION

A system for forming the image of a radar target is shown in Fig. 1. There are two modes for collecting data to form the image. An antenna of sufficient size may be used to focus radar energy

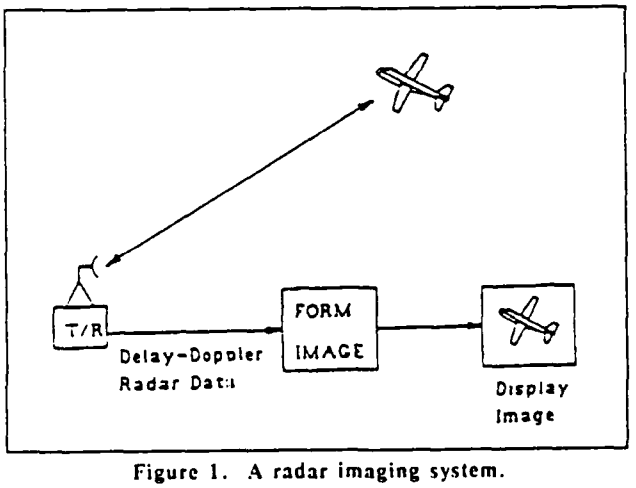


Figure 1. A radar imaging system.

onto a patch of the target having a size corresponding to a resolution element. By varying the position of the incident energy in some form of raster pattern over the target, all reflecting patches may be illuminated with a series of radar pulses and an image of the target formed by displaying the return energy for each patch. Alternatively, in spotlight mode, the energy is relatively unfocused, and the entire target is illuminated simultaneously by each transmitted pulse. The same form of image of energy versus range and cross-range coordinates can be formed from the more complicated echo data by suitable processing which utilizes delay and doppler-shift variations present in a series of target illuminations. Our concern is with forming high resolution images from data acquired in the spotlight mode.

There are at least two uses for high resolution images of a radar target. One is in developing a catalog of radar cross-section profiles for various target types. Another is for target identification. The latter use, illustrated in Fig. 2, normally proceeds in two separate and independent steps. First, the target's image is formed, and then features of the target, such as edges and textures, are extracted and used with any collateral information that may be available to identify the target. As illustrated in Fig. 3, we are also interested in developing a more coordinated approach to

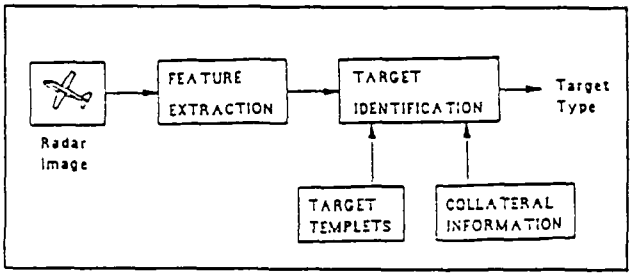


Figure 2. The use of an image for identification.

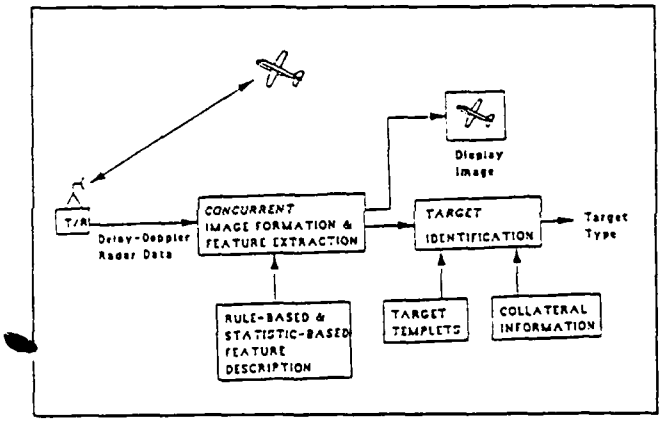


Figure 3. Coordinated target imaging and feature extraction.

forming an image of the target and extracting features so that these two processes may interact constructively so as to improve both. As a result of other developments in our laboratory [4,5], we believe that the method we have developed for forming target images is ideally suited for making this potentially important extension.

Wehner [1] and Mensa [2] describe methods for forming images of radar targets from spotlight mode data. A series of identical radar pulses (or pulse groups) having a linear FM chirp modulation is used to illuminate the target. The echoes are processed with a two-dimensional Fourier transform to form the image. This approach is based on an intuitive, deterministic analysis which results in accurate target images under high signal-to-noise ratio conditions. Our approach differs by incorporating a statistical model for the target's reflectivity, accommodating receiver noise, and in using a statistical estimation approach for developing a method for forming the image. A full development of our method is contained in [3].

TARGET MODEL

We model the complex envelope of the echo data according to:

$$r(t) = \int_{-\infty}^{\infty} s_T(t-\tau)\gamma(t-\tau,\tau)\,d\tau + w(t), \qquad (1)$$

where $w(t)$ is a white Gaussian noise with spectral density N_0, $s_T(t)$ is the complex envelope of the transmitted signal, and $\gamma(t,\tau)$ is the reflectivity at time t of all reflecting patches at two-way delay τ. $\gamma(t,\tau)$ may be expressed in terms of all reflectivities at

delay τ according to

$$y(t,\tau) = \int_{-\infty}^{\infty} c(f,\tau)e^{j2\pi ft}df. \quad (2)$$

where $c(f,\tau)$ is the reflectivity of all the points on the target at two-way delay τ which introduce a Doppler shift f. Since targets are of finite extent, both $y(t,\tau)$ and $c(f,\tau)$ are zero for τ outside some fixed interval, and $c(f,\tau)$ is zero for f outside a fixed interval because the Doppler variable is equivalent to the cross-range coordinate of a rotating target.

As we develop in detail in [3], (1) and (2) may be discretized into the matrix-vector form

$$\mathbf{r} = \Gamma^h \mathbf{c} + \mathbf{w}, \quad (3)$$

where the superscript "h" denotes the Hermitian transpose operation, $\mathbf{r}(l)$, $\mathbf{c}(l)$, and $\mathbf{w}(l)$ are vectors of samples of $r(t)$, $c(f,\tau)$, and $w(t)$, respectively, and Γ is a matrix each element of which is the product of a sample of $s_T(t)$ and a complex exponential.

We have adopted a *diffuse-target* model for the reflectivity. For this, the target is assumed to be comprised of uncorrelated scatterers each of which introduces a complex-valued, zero mean Gaussian random variable as a multiplicative factor on the incident signal reflected by it. The superposition of these according to (1) and (2) results in the radar echo data. This assumption implies that the reflectivity vector $\mathbf{c}$ has a Gaussian distribution with zero mean and diagonal covariance matrix $\Sigma = E(\mathbf{c}\mathbf{c}^h)$. This covariance consists of samples of the scattering function of the target, which is the power-spectrum of the reflectivity process $y(t,\tau)$. For our approach, the reflectivity is a two-dimensional Gaussian process, and the scattering function is its spectral intensity. Further, the received vector $\mathbf{r}$ has a Gaussian distribution with zero mean and covariance matrix

$$K_r = \Gamma^h \Sigma \Gamma + N_0 I. \quad (4)$$

The loglikelihood of the data $\mathbf{r}$ may be expressed in terms of this covariance matrix according to

$$L(K_r;\mathbf{r}) = -\ln(\det K_r) - \mathbf{r}^h K_r^{-1}\mathbf{r}. \quad (5)$$

IMAGING PROBLEM

Two different images of the target may be formed, one being an estimate of the target's reflectivity $\mathbf{c}$ and the other being an estimate of its scattering function Σ, both images being displayed in range and cross-range coordinates. A unique aspect of our method is that it produces a maximum-likelihood estimate of Σ and, also, a conditional mean estimate of $\mathbf{c}$ so that both of the possible images of the target can be displayed if desired.

Maximizing the loglikelihood (5) subject to the constraint that K_r must be of the parameterized form in (4) leads to a trace condition first discussed by Burg, Luenberger, and Wenger [6] and, in the present context, by Snyder, O'Sullivan, and Miller [3]. While this trace condition in principle specifies the maximum-likelihood estimate of the scattering function, it is a highly nonlinear equation with no closed form solution. For this reason, we have in [3] adopted the use of the alternating maximization approach of Dempster, Laird and Rubin [7] for producing the maximum-likelihood estimate numerically.

A sequence of estimates of the scattering function is obtained. The estimate at step $p+1$ is defined by the conditional expectation

$$\Sigma^{(p+1)} \overset{d}{=} E[\mathbf{c}\mathbf{c}^h | \Sigma^{(p)}, \mathbf{r}]. \quad (6)$$

where "$\overset{d}{=}$" means that the off diagonal terms on the left are zero and the diagonal terms equal the diagonal terms on the right. Evaluating the right side of (6) is a standard problem in estimation theory; it is given by

$$\Sigma^{(p)} + \Sigma^{(p)}\Gamma(\Gamma^h\Sigma^{(p)}\Gamma + N_0 I)^{-1}$$

$$\times(\mathbf{r}\mathbf{r}^h - \Gamma^h\Sigma^{(p)}\Gamma - N_0 I)(\Gamma^h\Sigma^{(p)}\Gamma + N_0 I)^{-1}\Gamma^h\Sigma^{(p)}. \quad (7)$$

Equations (6) and (7) define an iteration sequence which produces an estimate of the scattering function at each step. Thus, at each step, an image of the target may be displayed, and each image has successively higher likelihood. At each step, the conditional mean estimate of the reflectance is also generated. At step p, this estimate is

$$E[\mathbf{c}|\Sigma^{(p)},\mathbf{r}] = \Sigma^{(p)}\Gamma(\Gamma^h\Sigma^{(p)}\Gamma + N_0 I)^{-1}\mathbf{r}. \quad (8)$$

EXAMPLE

At the present time, we are implementing (6) and (7) to run on a Distributed Array Processor (DAP) made by the Active Memory Technology Company, which is a mesh connected parallel computer with 1024 processors. This implementation will permit comparisons to be made between this maximum-likelihood method and the more conventional method that employs two-dimensional Fourier transforms.

We have performed a preliminary computer-simulation study in which there is a point target that is concentrated at a single range and crossrange. For this situation, forming the target image corresponds to estimation of the power spectrum of a time series having one nonzero spectral component. Fig. 4 shows the result. The graph on the left in Fig. 4 is the output

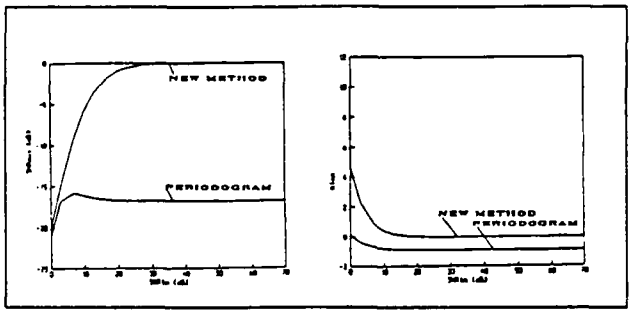


Figure 4. Shown are the output SNR (left) and bias (right) for estimating the reflected power of a point target in noise.

signal-to-noise ratio versus the input signal-to-noise ratio, and the graph on the right is bias versus the input signal-to-noise ratio, where these quantities were estimated from 3000 independent trials and are defined according to

$$\mathrm{SNR}_{\mathrm{in}} = E_0/N_0, \quad (9)$$

where E_0 is the average power in each of the P nonzero spectral components of the reflectivity, as defined by

$$E_0 = \frac{1}{P}\mathrm{Tr}(\Sigma);$$

and

$$\mathrm{SNR}_{\mathrm{out}} = \frac{\Sigma}{\sqrt{\mathrm{MSE}(\hat{\Sigma})}}, \quad (10)$$

where $\mathrm{MSE}(\hat{\Sigma})$ is the sample mean-square error in estimating Σ; and

$$\mathrm{BIAS} = \hat{\Sigma} - \Sigma. \quad (11)$$

The estimates of the scattering function for the maximum-likelihood method and for the two-dimensional Fourier transform method (which in one dimension becomes a periodogram) are compared in Fig. 4. The maximum-likelihood estimates (new method) are seen to have high output SNR and low bias compared to the periodogram for input SNRs above 5 dB. This very preliminary result gives us optimism that superior images of diffuse targets will be obtained with the maximum-likelihood method.

RELATED ISSUES

In [8], we have developed a Cramer-Rao lower bound on the mean-square error performance of maximum-likelihood estimates of the scattering function. The Fisher information matrix of this bound is shown in [8] to play an important role in specifying conditions for the uniqueness of estimates of the scattering function and in the selection of radar pulse shapes for achieving low variance images. Convergence of the iterative sequence in (6) and (7) is discussed in [9]. In [10], we develop a method for incorporating equality and inequality constraints on the input signal-to-noise ratio, which results in improved estimates when such side information is available.

ACKNOWLEDGEMENT

This work was supported by the Office of Naval Research under Contract Number N00014-86-K-0370 and by the National Science Foundation under Grant MIP-8722463.

REFERENCES

[1] D.R. Wehner, *High Resolution Radar*, Artech House, Dedham, MA, 1987.

[2] D.L. Mensa, *High Resolution Radar Imaging*, Artech House, Dedham, MA, 1984.

[3] D.L. Snyder, J.A. O'Sullivan, and M.I. Miller, "The Use of Maximum-Likelihood Estimation for Forming Images of Diffuse Radar Targets from Delay-Doppler Data," *IEEE Trans. on Information Theory*, to appear in 1989.

[4] M.I. Miller, B. Roysam, K. Smith, and J. T. Udding, "On the Equivalence of Regular Grammers and Stochastic Constraints: Applications to Image Processing on Massively Parallel Processors," *AMS-IMS-SIAM Joint Conf. on Spatial Statistics and Imaging*, American Math. Society, July 1988.

[5] M.I. Miller, B. Roysam, and K. Smith, "Mapping Rule-Based and Stochastic Constraints to Connection Architectures: Implication for Hierarchical Image Processing," *Proc. SPIE*, Cambridge, MA, Nov. 1988.

[6] J.P. Burg, D.G. Luenberger, and D.L. Wenger, "Estimation of Structured Covariance Matrices," *Proc. IEEE*, Vol. 70, pp. 963-974, Sep't. 1982.

[7] A.D. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Society*, Vol. B39, pp. 1-37, 1977.

[8] J.A. O'Sullivan, P. Moulin, and D.L. Snyder, "Cramer-Rao Bounds for Constrained Spectrum Estimation with Application to a Problem in Radar Imaging," *Proc. 1988 Allerton Conf.*, Univ. of Illinois, Urbana, 1988.

[9] J.A. O'Sullivan and D.L. Snyder, "High Resolution Radar Imaging using Spectrum Estimation Methods," *Proc. Workshop on Signal Processing*, Institute for Mathematics and Its Applications, University of Minnesota, Minneapolis, MN, 1988.

[10] P. Moulin, D.L. Snyder, and J.A. O'Sullivan, "Constrained Maximum Likelihood Spectrum Estimation of Periodic Processes from Noisy Data," submitted for presentation at the 1989 Conf. on Information Sciences and Systems, The Johns Hopkins Univ., Baltimore, MD.

6.4 Reprint of: P. Moulin, D. L. Snyder, and J. A. O'Sullivan: "A Sieve-Constrained Maximum-Likelihood Estimator for the Spectrum of a Gaussian Process", Proc. 28th Allerton Conference, Urbana-Champaign, IL, Sept. 1989.

A SIEVE-CONSTRAINED MAXIMUM-LIKELIHOOD ESTIMATOR FOR THE SPECTRUM OF A GAUSSIAN PROCESS *

P. Moulin, D.L. Snyder, J.A. O'Sullivan
Electronic Systems and Signals Research Laboratory
Department of Electrical Engineering
Washington University
Saint Louis, MO 63130

ABSTRACT

Maximum-likelihood spectrum estimation is an ill-posed problem. In this paper, we use a method of sieves for addressing this issue. The estimate of the spectrum is constrained to a subset of some Hilbert space of functions over which a complete set of nonorthogonal basis functions is defined. The estimate is then represented by a countable set of coefficients in a nonorthogonal series expansion. By defining an appropriate sieve on this countable set, our problem reduces to maximum-likelihood estimation of the parameters in the sieve. Three main attractive features of this approach are: (1) the nonorthogonal expansion is a convenient framework for defining the sieve and including *a priori* information; (2) mean-square consistency of the estimates can be expected; and (3) we have derived a tractable alternating maximization algorithm for estimating the parameters. The setup of this problem is general and can be applied without major difficulties to the estimation of higher-dimensional spectral functions, as occurs, for example, in imaging radar targets from delay-doppler data.

1. INTRODUCTION

Maximum-likelihood (ML) estimation of a continuous function, or even of an infinitely countable set of parameters is known to be an ill-posed problem [1]. In this paper, we use a regularization method for addressing this issue when the continuous function is the spectral density of a Gaussian process, and this density must be estimated from one single sample function of the process. The ill-posedness of the estimation problem is often addressed by approximating the model for the process by a simpler one, assuming temporal periodicity. The spectral density is then discrete. It follows that for bandlimited processes, the spectrum is finite-dimensional and can be estimated by means of standard ML techniques. However, the fundamental difficulties inherent to ill-posed problems remain, indicating that the assumption of a periodic process is not sufficient. For instance, the length of the periodic extension of the data cannot be made arbitrarily large without the estimates becoming very rough. Furthermore, the estimate obtained from the observation of one single realization of the process is not consistent, no matter how many samples are collected.

The approach we use for estimating the continuous spectral density does not require any approximation of the model. It is based upon Grenander's method of sieves, which provides a framework for estimation in an infinite-dimensional parameter space [1]. The estimation is performed in a restricted parameter set over which the estimates are stable. For an infinite amount of data, this restricted parameter set is dense in the actual parameter set. This procedure leads to producing stable and, we expect, consistent estimates.

Section 2 of this paper contains a description of the model for the stochastic process whose spectral density is sought. In Section 3, we briefly summarize some previous results in the literature, in which an approximation of the model is made; the limitations of these methods are mentioned. Section 4 contains a short review of regularization methods. The sieve-constrained ML estimator is described in Section 5.

* The work described in this paper was supported by contract number N00014-86-K-0370 from the Office of Naval Research and by grant number MIP-8722463 from the National Science Foundation.

2. MATHEMATICAL MODEL

Consider a complex-valued, wide-sense stationary stochastic-process $y(t)$. Our estimation procedure is based upon the following statistical model for the observations. The process is Gaussian with a mean of zero. Its spectrum $S(f)$ is bandlimited to some f_{max}, and N samples of one realization of the process are observed, with the sampling frequency $1/\Delta t$ at least equal to the Nyquist rate $2f_{max}$. Furthermore, the process is observed in complex-valued additive white Gaussian noise $w(t)$ with spectral density N_0.

Under the assumptions stated above, the samples of $y(t)$ can be represented via the Cramér spectral representation:

$$y(n\Delta t) = \int_{-f_{max}}^{f_{max}} \exp(-j2\pi f n\Delta t)\, Z(df) \qquad , 0 \le n \le N-1 . \tag{2.1}$$

In this generalized Stieltjes integral, $Z(df)$ is an orthogonal, Gaussian spectral process with variance $S(f)\,df$ [1, p.74].

In the next sections, a ML estimator for $S(f)$ is presented. We shall find it convenient to consider a discrete approximation to the integral (2.1). Define a uniform partitioning of the interval $[-f_{max}, f_{max}]$, $\{f_k\}_{k=0}^{M-1}$, with partition size $\Delta_M = 2f_{max}/M$. Next define the process

$$c(k) = (\Delta_M)^{-1/2}\,[Z(f_{k+1}) - Z(f_k)] \qquad , 0 \le k \le M-1 . \tag{2.2}$$

This process is Gaussian and orthogonal. The variance of each sample,

$$\sigma^2(k) = \frac{1}{\Delta_M} \int_{f_k}^{f_{k+1}} S(f)\, df \ , \tag{2.3}$$

is an approximation to $S(f_k)$. Next, approximating the Stieltjes integral (2.1) with a Riemann sum, we get

$$\tilde{y}(n\Delta t) = \sqrt{\frac{2f_{max}}{M}} \sum_{k=0}^{M-1} \exp(-j2\pi f_k n\Delta t)\, c(k) . \tag{2.4}$$

As the partition size tends to zero, $\tilde{y}(.)$ converges to the stochastic integral $y(.)$ in the mean-square sense:

$$y(n\Delta t) = \underset{M\to\infty}{l.i.m.}\ \tilde{y}(n\Delta t) \tag{2.5}$$

Notice that in the special case of a periodic process with period $M\Delta t$, $\tilde{y}(.) = y(.)$, so that (2.4) is the *exact* spectral representation of $y(n\Delta t)$. This observation has motivated using (2.4) as a model for $y(.)$ in [2]. The parameter M is greater or equal to N but is not further specified, but M should be large enough to ensure that the model approximation is valid. In [2] the issue of ill-posedness is addressed by keeping M finite. In the next section, we will summarize some known results for this estimator. On the other hand, the sieve-constrained ML estimator presented here is still based on the representation (2.4), but does not require M to be finite. As such, this estimator does not involve any model approximation, as shown in (2.5).

We now specify the model for the observations. Using the same vector notations as in [2], we get, from (2.4):

$$r = \Gamma^{\dagger} c + w \ , \tag{2.6}$$

where r is the N-vector of observations, w is a noise vector with diagonal covariance $N_0 I_N$, where I_N is the $N\times N$ identity matrix, and c is the M-vector formed from the samples of the orthogonal process $c(k)$ defined in (2.2). Since c is orthogonal, its covariance Σ is diagonal with entries $\{\sigma^2(k)\}_{k=0}^{M-1}$. Γ is a $M\times N$ matrix with (k,n) entry given by $\sqrt{\frac{2f_{max}}{M}}\exp(-j2\pi f_k n\Delta t)$; finally, $\dagger$ denotes the Hermitian-transpose operator for matrices *.

* Γ can be generalized readily to accomodate a linear transformation of $y(.)$ in the measurement, as occurs in the radar

The covariance matrix for r is given by

$$K_r = E[rr^\dagger] = \Gamma^\dagger \Sigma \Gamma + N_0 I_N \,. \tag{2.7}$$

3. UNCONSTRAINED ML ESTIMATOR

In this section, we introduce the ML estimator for the spectral matrix Σ defined in the model (2.6). From (2.7), the loglikelihood function for Σ is

$$L(r,\Sigma) = -\ln\det(\Gamma^\dagger \Sigma \Gamma + N_0 I_N) - r^\dagger (\Gamma^\dagger \Sigma \Gamma + N_0 I_N)^{-1} r \,. \tag{3.1}$$

Maximizing the likelihood with respect to Σ yields the necessary trace condition which a positive definite estimate $\hat{\Sigma}$ must satisfy [2]:

$$Tr\ [\Gamma \hat{K}_r^{-1}(rr^\dagger - \hat{K}_r)\hat{K}_r^{-1}\Gamma^\dagger \delta\hat{\Sigma}] = 0 \,, \tag{3.2}$$

with

$$\hat{K}_r = \Gamma^\dagger \hat{\Sigma} \Gamma + N_0 I_N \,,$$

for all admissible $M \times M$ diagonal matrices $\delta\hat{\Sigma}$ [2]. This trace condition is a nonlinear equation in $\hat{\Sigma}$. Generally it cannot be solved directly in closed-form, so some numerical search procedure must be implemented. An approach is the expectation-maximization (EM) algorithm of Dempster *et al.* [3] used in [2]. The complete data are defined as (c,w). An initial estimate $\hat{\Sigma}^{(0)}$ is selected. At step k+1 (k = 0,1,..) the estimate is updated according to

$$\hat{\Sigma}^{(k+1)} = \underset{\Sigma}{argmax}\ Q(\Sigma|\hat{\Sigma}^{(k)}) \tag{3.3}$$

where

$$Q(\Sigma|\hat{\Sigma}^{(k)}) = -\sum_{i=0}^{M-1} \ln \sigma^2(i) - \sum_{i=0}^{M-1} \frac{E[|c(i)|^2|r,\hat{\Sigma}^{(k)}]}{\sigma^2(i)} \,, \tag{3.4}$$

and

$$E[|c(i)|^2|r,\hat{\Sigma}^{(k)}] = [\hat{\Sigma}^{(k)} - \hat{\Sigma}^{(k)}\Gamma K_r^{-1(k)}\Gamma^\dagger \hat{\Sigma}^{(k)} + \hat{\Sigma}^{(k)}\Gamma K_r^{-1(k)} rr^\dagger K_r^{-1(k)}\Gamma^\dagger \hat{\Sigma}^{(k)}]\,(i,i) \,, \tag{3.5}$$

where

$$K_r^{(k)} = \Gamma^\dagger \hat{\Sigma}^{(k)} \Gamma + N_0 I_N \,.$$

This algorithm produces a sequence of estimates

$$\hat{\sigma}^2(i)^{(k+1)} = E[|c(i)|^2|r,\hat{\Sigma}^{(k)}] \tag{3.6}$$

having increasing likelihood. It can be shown that the stable points of this algorithm satisfy the necessary trace condition (3.2) for a maximizer [2].

The performance of this estimator has been evaluated in previous studies and compared to traditional spectrum estimators such as the periodogram [4, 5]. It has been shown that

(1) the ML estimates have a very low bias;

(2) the ML estimates offer a better trade-off between variance and resolution than the periodogram;

(3) the ML estimates are not mean-square consistent. This unfortunate result is due to the fact that the dimension M of the parameter space is at least equal to the number of data N. We are actually dealing with a *small-sample* ML estimation problem. This observation is one factor that motivated the study presented in the next sections.

imaging problem [2].

4. REGULARIZATION OF THE ML ESTIMATES

According to Hadamard's classical definition, a problem is said to be well posed if: (1) it has a solution, (2) the solution is unique, and (3) the solution varies continuously with the data. A problem is ill posed if it is not well posed. Small changes in the data can then produce unbounded changes in the estimates. Typical examples are inversion of certain Fredholm integral equations and the reconstruction of functions from truncated Fourier transforms [6].

The ML estimation procedure produces similar artifacts when the number of parameters (M) is large compared to the number of data samples (N). The estimates are extremely sensitive to small changes in the data and exhibit a very rough shape characterized by very sharp peaks and low valleys. We can say that the problem is *practically*, if not *technically*, ill posed. Such a behavior is highly undesirable, and a *regularization* of the estimates is required. This concept was introduced by Tikhonov [7]. A large bulk of mathematical literature has been written on this subject; see for instance Bertero [8] for a tutorial.

One possible method for the regularization of estimates is the method of sieves, introduced by Grenander [1]. According to Grenander's definition, a *sieve in a parameter space A* is a family of subsets $S(\mu)$ of A indexed by a positive parameter μ called the *mesh size*. A restricted ML estimate exists over each set $S(\mu)$. As the mesh size tends to zero the sets $S(\mu)$ will be large enough to allow the ML solution to converge to any solution in A [1, p.357]. For the problem at hand, the choice of a "good" value of μ will depend on the data record size, the noise level, and possibly other factors. A major problem is to find sieves which will make the ML estimates converge and possess some desirable practical features such as analytical and/or computational tractability. For instance, one might consider a *convolution sieve:*

$$S(\mu) = \left\{ \sigma^2(i) \mid \sigma^2(i) = \eta(i) * \psi_\mu(i) \;\; 0 \le i \le M-1 \right\}, \tag{4.1}$$

where $\sigma^2(i)$ is the constrained spectrum, $\eta(i)$ is any spectrum, and $\psi_\mu(i)$ is a convolution kernel, or "window", with mesh equal to μ. However, it is not clear that finding spectral estimates within this sieve is computationally tractable. In 1985 Chow and Grenander recognized that regularizing the ML spectrum estimates with a convolution sieve appears to be a *formidable* task, so they studied other types of estimators [9].

Here we recommend a regularization method based on nonorthogonal expansions of the unknown density function. This approach leads to computationally tractable ML estimates. As discussed in §5.3, we also expect the estimates to be asymptotically consistent.

The regularization procedure developed in the next section will be applied to the technically ill-posed spectrum estimation problem in which the ratio M/N tends to infinity, where the approximate model (2.4) converged to the exact model (2.1). We will show how to solve this infinite-dimensional parameter estimation problem.

5. A METHOD OF SIEVES BASED ON NONORTHOGONAL EXPANSIONS

The spectrum $S(f)$ belongs to a Hilbert space of functions defined over $[-f_{max}, f_{max}]$. Denote by $\mathbf{H}$ this Hilbert space, and by $\mathbf{H}_+$ the subset of all nonnegative functions in $\mathbf{H}$. Typically $\mathbf{H}$ is $L_2[-f_{max}, f_{max}]$, the space of square-integrable functions over $[-f_{max}, f_{max}]$. Our goal is to represent the spectral function in $\mathbf{H}_+$ in terms of a set of basis functions having localized support. This permits a good *local* representation of a function. As such, this is related to the concept of wavelets.

5.1. Definition of the sieve

Consider a spectral M-vector $\underline{\sigma}^2$. To this vector we associate a step function $S(f)$ according to the map ST_M defined as follows:

$$ST_M[\underline{\sigma}^2] : \; \mathbf{R}^M \to \mathbf{H} : \; \underline{\sigma}^2 \to S(f) = ST_M[\underline{\sigma}^2] = \sum_{i=0}^{M-1} \underline{\sigma}^2(i)\, \chi_i(f), \tag{5.1}$$

where $\{f_i\}_{i=0}^{M-1}$ is the same uniform partition of the interval $[-f_{max}, f_{max}]$ as in Section 2, R^M is the M-dimensional Euclidean space, and $\chi_i(f)$ is the indicator function over the i-th partition interval.

Let $\{\psi_m(f)\}_{m=0}^{\infty}$ be a nonorthogonal set of positive basis functions in H_+. Consider the set A of all functions in the span of the $\{\psi_m\}_{m=0}^{\infty}$, with positive coefficients:

$$A = \left\{ S(f) \mid S(f) = \sum_{m=0}^{\infty} a(m)\psi_m(f) \quad , a(m) > 0 \right\}. \tag{5.2}$$

We define our parameter set to be A. Clearly A is a subset of H_+. It would be desirable to have $\{\psi_m\}_{m=0}^{\infty}$ be designed so that A is dense in H_+. It is well known that this can be achieved with $\{\psi_m\}_{m=0}^{\infty}$ defined as indicator functions over a partition of the frequency domain, *in the limit as the partition size tends to zero.* How A is made dense in H_+ for more elaborate designs remains to be investigated.

Now define the following step-function approximation of $\psi_m(f)$:

$$\psi_{M,m}(f) = ST_M[\underline{\psi}_{M,m}], \tag{5.3}$$

where $\underline{\psi}_{M,m}$ is a M-vector of samples of $\psi_m(f)$ taken on the partition of the frequency interval. As $M \to \infty$, $\psi_{M,m}(f)$ converges to $\psi_m(f)$ in the norm of H. We define a sieve $S(M,Q)$ on $S(f)$ in A by truncating the series representation (5.2) to a finite number of terms Q and considering M-step approximations to the basis functions,

$$S(M,Q) = \left\{ S(f) \mid S(f) = \sum_{m=0}^{Q-1} a(m)\psi_{M,m}(f) \quad , a(m) > 0 \right\}. \tag{5.4}$$

The functions in the set $S(M,Q)$ are constrained to be M-step functions of the type (5.1). For the time being, we keep M finite, but ultimately M will be allowed to tend to infinity. We view $\{a(m)\}_{m=0}^{Q-1}$ as being Q unknown parameters for which ML estimates are sought.

Why nonorthogonal basis functions?

Before considering the general problem, we mention the special case where the basis functions considered in (5.4) are Q indicator functions over disjoint intervals covering the frequency domain. In this instance, the computation of the $a(m)$'s is straightforward *. However these basis functions have the major disadvantage that the estimates are step functions, and so display discontinuities that are uncharacteristic of spectral densities.

To achieve good representations for a richer class of functions, we have to forgo the aforementioned restriction on the structure of the basis functions. A possible design of the basis functions might be based upon *dilations* and *translations* of a smooth basic function. Such a design is used in the wavelet literature because it provides a convenient *multiscale* representation of the function of interest [10, 11]. This is usually achieved more easily with nonorthogonal basis functions [11].

Our strategy in using other basis functions will have practical interest if a tractable parameter estimation procedure can be derived. This issue is addressed in the next section.

5.2 Computation of the sieve estimator

To each function $S(f)$ in the constrained set (5.4) is associated a spectral M-vector $\underline{\sigma}^2$ via the inverse map

$$\underline{\sigma}^2 = ST_M^{-1}[S(f)]. \tag{5.5}$$

* The maximization problem (5.7) can then be solved in closed form.

Since the map ST_M^{-1} is linear, we conclude from (5.1), (5.4) and (5.5) that the vector $\underline{\sigma}^2$ is constrained as follows.

$$\underline{\sigma}^2 = \sum_{m=0}^{Q-1} a(m)\underline{\psi}_{M,m} \,. \tag{5.6}$$

The ML estimates of the $a(m)$'s can be obtained by using the EM algorithm described in Section 3. At each step the expression (3.4) is maximized with respect to $\hat{\sigma}^2$ subject to the sieve constraint (5.5). This amounts to finding the maximum with respect to a of

$$Q(a|\hat{a}^{(k)}) = -\sum_{i=0}^{M-1} \ln\left[\sum_{m=0}^{Q-1} a(m)\underline{\psi}_{M,m}(i)\right] - \sum_{i=0}^{M-1} \frac{E[|c(i)|^2|r,\hat{a}^{(k)}]}{\sum_{m=0}^{Q-1} a(m)\underline{\psi}_{M,m}(i)} \,. \tag{5.7}$$

This maximization problem is very difficult and must be solved via some numerical procedure. Keeping in mind that this operation has to be repeated at every iteration of the EM-algorithm, one is hardly attracted by this prospect. We propose the following method instead.

An EM algorithm has been developed, based on the definition of a new complete data space for producing the estimates. This method is set up as an unconstrained maximization problem in a Q-dimensional parameter space and, as such, it does not involve any computation in **H**. Because of the existence of such an algorithm, our estimation procedure based on nonorthogonal expansions turns out to be computationally feasible. This usually represents an impressive computational saving and is a major attractive feature of this approach.

Complete/Incomplete Data Spaces

Write c as the sum of Q independent vectors:

$$c = \sum_{m=0}^{Q-1} \tilde{c}_m \,, \tag{5.8}$$

where $\tilde{c}_m$ is a M-vector, sample of a 0-mean Gaussian process with diagonal covariance $a(m)\tilde{\Psi}_m$, with $\tilde{\Psi}_m$ a diagonal matrix made of $\{\ \underline{\psi}_{M,m}(i), i \in [0..M-1]\ \}$. These Q processes are independent. Now define the J_m-vector c_m as the restriction of $\tilde{c}_m$ to its support set S_m. The covariance Ψ_m of c_m is made of the nonzero entries of $\tilde{\Psi}_m$. With these notations our model (2.6) becomes

$$\begin{aligned} r &= \sum_{m=0}^{Q-1} \Gamma^t \tilde{c}_m + w \\ &= \sum_{m=0}^{Q-1} \Gamma_m^t c_m + w \,. \end{aligned} \tag{5.9}$$

We define the complete data as ($\{c_m\}_{m=0}^{Q-1}$; w). Following this definition we write the loglikelihood for the complete data as

$$\begin{aligned} l_{cd}(a) &= \sum_{m=0}^{Q-1} \ln p(c_m : a) \\ &= -\sum_{m=0}^{Q-1} \ln\det(a(m)\Psi_m) - \sum_{m=0}^{Q-1} c_m^t (a(m)\Psi_m)^{-1} c_m \,. \end{aligned}$$

Discarding all terms not involving a, we obtain:

$$l_{cd}(a) = -\sum_{m=0}^{Q-1} J_m \ln a(m) - \sum_{m=0}^{Q-1} \frac{1}{a(m)} \sum_{i \in S_m} \frac{|c_m(i)|^2}{\underline{\psi}_{M,m}(i)} \,. \tag{5.10}$$

The conditional expectation of $l_{cd}(a)$ is then

$$Q(a|\hat{a}^{(k)}) = -\sum_{m=0}^{Q-1} J_m \ln a(m) - \sum_{m=0}^{Q-1} \frac{1}{a(m)} \sum_{i \in S_m} \frac{E[|c_m(i)|^2|r,\hat{a}^{(k)}]}{\underline{\psi}_{M,m}(i)} \,. \tag{5.11}$$

Taking the partial derivative of (5.11) with respect to $a(m)$, and setting the result to zero yields

$$\hat{a}(m)^{(k+1)} = \frac{1}{J_m} \sum_{i \in S_m} \frac{E\,[|c_m(i)|^2 | r, \hat{a}^{(k)}]}{\underline{\Psi}_{M,m}(i)} . \tag{5.12}$$

Next we evaluate the conditional expectation of $|c_m(i)|^2$ in (5.12). Define

$$\hat{c}_m^{(k)} = E\,[\, c_m \,|\, r, \hat{a}^{(k)}] . \tag{5.13}$$

The conditional expectation in (5.12) is the i-th diagonal element of the matrix

$$\begin{aligned} E\,[\, c_m c_m^\dagger | r, \hat{a}^{(k)}] &= E\,[\,(c_m - \hat{c}_m^{(k)})\,(c_m - \hat{c}_m^{(k)})^\dagger | r, \hat{a}^{(k)}] + \hat{c}_m'^{(k)} \hat{c}_m^{(k)\dagger} \\ &= (K_{c_m r}^{(k)} K_r^{-1(k)} r)(K_{c_m r}^{(k)} K_r^{-1(k)} r)^\dagger + K_{c_m c_m}^{(k)} - K_{c_m r}^{(k)} K_r^{-1(k)} K_{r c_m}^{(k)} , \end{aligned} \tag{5.14}$$

where we denote by K_{xy} the conditional correlation $E\,[xy^\dagger | \hat{a}^{(k)}]$ of two random vectors x and y. Now, the expectations are evaluated from the model (5.9). Note that because the c_m's are independent, the conditional expectations on $\hat{a}^{(k)}$ are simply conditional expectations on $\hat{a}(m)^{(k)}$. We get

$$K_{c_m r}^{(k)} = E\,[\, c_m r^\dagger | \hat{a}^{(k)}] = \hat{a}(m)^{(k)} \Psi_m \Gamma_m , \tag{5.15a}$$

$$K_r^{(k)} = E\,[\, r r^\dagger | \hat{a}^{(k)}] = \sum_{j=0}^{Q-1} \hat{a}(j)^{(k)} \Gamma_j^\dagger \Psi_j \Gamma_j + N_0 I_N , \tag{5.15b}$$

$$K_{c_m c_m}^{(k)} = E\,[\, c_m c_m^\dagger | \hat{a}^{(k)}] = \hat{a}(m)^{(k)} \Psi_m . \tag{5.15c}$$

Substituting the expressions (5.15a,c) into (5.14) yields

$$E\,[\, c_m c_m^\dagger | r, \hat{a}^{(k)}] = \hat{a}(m)^{(k)} \Psi_m + \hat{a}(m)^{(k)} \Psi_m \Gamma_m K_r^{-1(k)} [\, r r^\dagger - K_r^{(k)}] K_r^{-1(k)} \Gamma_m^\dagger \Psi_m \hat{a}(m)^{(k)} , \tag{5.16}$$

with $K_r^{(k)}$ given in (5.15b). Taking the i-th diagonal element of (5.16) and substituting in (5.12), we get

$$\begin{aligned} \hat{a}(m)^{(k+1)} &= \frac{1}{J_m} \sum_{i \in S_m} \frac{1}{\underline{\Psi}_{M,m}(i)} \Big[\hat{a}(m)^{(k)} \underline{\Psi}_{M,m}(i) + \hat{a}(m)^{(k)} \underline{\Psi}_{M,m}(i) \\ &\qquad \times [\Gamma_m K_r^{-1(k)} (r r^\dagger - K_r^{(k)}) K_r^{-1(k)} \Gamma_m^\dagger]_{ii} \underline{\Psi}_{M,m}(i) \hat{a}(m)^{(k)} \Big] \\ &= \hat{a}(m)^{(k)} + \hat{a}(m)^{(k)} \Bigg[\frac{1}{J_m} \sum_{i \in S_m} \underline{\Psi}_{M,m}(i) \\ &\qquad \times [\Gamma_m K_r^{-1(k)} (r r^\dagger - K_r^{(k)}) K_r^{-1(k)} \Gamma_m^\dagger]_{ii} \Bigg] \hat{a}(m)^{(k)} . \end{aligned} \tag{5.17}$$

Now defining the $N \times N$ matrices

$$Z^{(k)} = K_r^{-1(k)} (r r^\dagger - K_r^{(k)}) K_r^{-1(k)} \tag{5.18}$$

and

$$K_m = \frac{1}{J_m} \Gamma_m^\dagger \Psi_m \Gamma_m , \tag{5.19}$$

the term between brackets in the right-hand side of (5.17) can be written

$$\begin{aligned} \frac{1}{J_m} \sum_{i \in S_m} \underline{\Psi}_{M,m}(i) (\Gamma_m Z^{(k)} \Gamma_m^\dagger)_{ii} &= \frac{1}{J_m} Tr \Big[\Psi_m \Gamma_m Z^{(k)} \Gamma_m^\dagger \Big] \\ &= Tr \Big[\frac{1}{J_m} \Gamma_m^\dagger \Psi_m \Gamma_m Z^{(k)} \Big] \\ &= Tr \Big[K_m Z^{(k)} \Big] . \end{aligned} \tag{5.20}$$

Substituting (5.20) in (5.17) we get the update equations at stage k of the algorithm:

$$\hat{a}(m)^{(k+1)} = \hat{a}(m)^{(k)} + \hat{a}(m)^{(k)} Tr \Big[K_m Z^{(k)} \Big] \hat{a}(m)^{(k)} , \tag{5.21}$$

$$K_r^{(k+1)} = \sum_{j=0}^{Q-1} \hat{a}(j)^{(k)} K_j + N_0 I_N \ . \qquad (5.22)$$

Comments

(1) The update equations (5.21) and (5.22) highlight the role played by the basis covariances K_m in this representation.

(2) As $M \to \infty$, the algorithm for estimating the coefficients $a(m)$ is well-behaved. The reason is that the basis covariance K_m in (5.21) is then given by the limit of the matrix product (5.19) as $M \to \infty$, which is a matrix of integrals with entry (n,m) given by

$$K_m(n,m) = \int_{-f_m}^{f_m} \exp(-j2\pi f(n-m)\Delta t)\, \psi_m(f)\, df \ .$$

We can thus let $M \to \infty$ without any loss in stability of the estimates or any increse in computations.

(3) Implementation of the solution:

Each matrix K_m is constant and can be computed off-line. At each step the covariance matrix K_r in (5.22) must first be inverted, which accounts for N^3 operations. Following this, updating each $\hat{a}(m)$ just requires N^2 multiplications/additions. The complexity of each step of the EM algorithm is then $N^3 + QN^2$, as compared to $N^3 + MN^2$ in the unconstrained case. The saving is impressive for infinite M, keeping $Q \sim N$.

(4) It is easily shown that the trace appearing in the update equation (5.21) is the partial derivative of the incomplete-data loglikelihood with respect to $a(m)$.

5.3. Some Regularization Aspects

The most attractive feature of this approach is a theoretical one. We are effectively estimating Q parameters instead of M as in the unconstrained problem. The choice of Q is somewhat arbitrary and depends on the way we have decided to represent the spectral function. As a rule of thumb it seems desirable to have Q (# of parameters) no larger than N (# of data). This can be done while letting M be infinite, a result that could not have been obtained without regularization of the ML estimates. The process in our model then approaches the limiting case of a (nonperiodic) stationary process, and as such does not involve any model approximation.

Consistency of the estimates

A major goal when defining a sieve is to ensure consistency of the estimates. Although no thorough study has yet been undertaken for the sieve (5.4), we believe that consistency can be achieved, provided that Q tends to infinity at a slower rate than N. Typically : $Q \sim N^{\alpha}$, with $0 < \alpha < 1$. Another objective is to ensure rapid convergence of the estimates. Clearly the issue here is the design of the basis functions, which is one of our current research areas.

CONCLUSIONS

We have proposed a method for a estimating a spectral function represented by a linear combination of basis functions. The task of finding the ML estimates of the coefficients of these basis functions looks formidable at first (refer to the optimization problem (5.7)). However we propose a method based on the decomposition (5.8) of the spectral process into a sum of Q independent spectral processes with known covariance, up to a scale factor. The subsequent reformulation of the complete/incomplete data spaces leads to the derivation of a powerful algorithm for evaluating these scale factors.

The main features of this approach are: (1) *flexibility*. The choice of basis functions is wide open. In particular we can choose the shape of the basis functions, amount of overlap, and support set extent (possibly non-uniform). (2) *using an exact model*. No approximation of the exact model for the stochastic process is needed. (3) *computational efficiency*. The computational complexity of the algorithm is a function of the dimension of the parameter space, not of the partition size of the frequency axis. An important result is that it is now possible to let $M \to \infty$ without any increase in the number of computations

Several interesting questions remain open issues. In particular it is desired to achieve mean-square consistency of the estimates. As mentioned in §5.2.2, this might be done by letting the sieve grow at an appropriate rate as $N \to \infty$. The issue of uniqueness of the estimates should also be investigated. Finally, the design of the basis functions will determine the overall estimator performance, as far as convergence rate and sensitivity to noise are concerned.

Also note that the setup of this problem is very general and can be applied without major difficulties to the estimation of higher-dimensional spectral functions. In particular, the spectra of interest in the radar-imaging problem described in [2] are two-dimensional. The application of the ideas in this paper to the radar-imaging problem is another area of active research.

REFERENCES

[1] U. Grenander, *Abstract Inference*, Wiley 1981.

[2] D. L. Snyder, J. A. O'Sullivan and M. I. Miller, "The Use of Maximum-Likelihood Estimation for Forming Images of Diffuse Radar-Targets from Delay-Doppler Data", Vol. 35, No. 3, *IEEE Trans. on Information Theory*, pp. 536-548, 1989.

[3] A.D. Dempster, N.M. Laird and D.B. Rubin, "Maximum-Likelihood from Incomplete Data via the EM Algorithm", *J. Royal Stat. Soc., Vol. B39*, pp. 1-37, 1977.

[4] P. Moulin, "Maximum-Likelihood Spectrum Estimation of Periodic Processes from Noisy Data", *Washington University, ESSRL internal report.*

[5] P. Moulin, D.L. Snyder and J.A. O'Sullivan, "Maximum-Likelihood Spectrum Estimation of Periodic Processes from Noisy Data", *Proc. of the 1989 Conference on Information and Systems Science*, Johns Hopkins University, Baltimore, 1989.

[6] L.S. Joyce and W.L. Root, "Precision Bounds in Superresolution Processing", *J. Opt. Soc. Am., Vol.1*, pp. 149-168, 1984.

[7] A. Tikhonov and V. Arsenin, *Solutions of Ill-Posed Problems*, Winston & Sons, Washington D.C. 1977.

[8] M. Bertero, C. De Mol, G. Viano, "The Stability of Inverse Problems", Chapter 5 in *Inverse Scattering Problems*, Springer-Verlag 1980.

[9] Y. Chow and U. Grenander, "A Sieve Method for the Spectral Density", *Ann. Stat., Vol. 13, No. 3*, pp. 998-1010, 1985.

[10] M. Frasier and B. Jawerth, "Decomposition of Besov Spaces", *Indiana Univ. Math. Journal, Vol. 34, No. 3*, 1985, pp.777-799.

[11] I. Daubechies, A. Grossmann and Y. Meyer, "Painless Nonorthogonal Expansions", *J. Math. Phys., Vol. 27, No. 5*, 1986.

6.5 Reprint of: J. A. O'Sullivan, P. Moulin, D. L. Snyder, and S. P. Jacobs, "Computational Considerations for Maximum-Likelihood Radar Imaging," Proc. 1990 CISS, Princeton University.

Computational Considerations for Maximum-Likelihood Radar Imaging†

J. A. O'Sullivan
P. Moulin
D. L. Snyder
S. P. Jacobs

Electronic Systems and Signals Research Laboratory
Department of Electrical Engineering
Washington University
Saint Louis, MO 63130

ABSTRACT

Recent papers have outlined a new approach for spectrum estimation and radar imaging based on expectation-maximization algorithms for structured covariance estimation. Performance of this approach has been promising for the problems studied. Application of this approach to real data sets has been limited, however, due to the need to invert a matrix whose dimension equals the size of the data set. For radar applications where an image is to be formed, data sets can be on the order of 2^{14} for 128×128 images. This makes the use of the new approach difficult in its previously described form. This paper proposes both approximation methods for inverting typical matrices and constraints on radar transmitted signals which make maximum likelihood image estimation viable. These constraints may be satisfied for real signals used in radar imaging systems. Simulations are shown to demonstrate the performance of the algorithms. Finally, motivated by the images resulting from the simulations, regularization methods are discussed.

Introduction

New approaches are being studied for maximum likelihood spectrum estimation and radar imaging which are based on using the expectation-maximization algorithm [1] for structured covariance estimation [2-4]. In this paper, we focus on the radar imaging problem, although many of the results hold for similar spectrum estimation problems. The limitations of our previous algorithm [4] are very clear for the radar imaging problem with large data sets. In order to form an image from a data set of size N, a matrix of dimension $N \times N$ must be inverted. When N is on the order of 2^{14}, this inversion is not practical without exploiting its special structure or making some approximations.

First, the equations which describe the radar data are defined. Next, the algorithm derived in [4] is presented. After discussing the role of the matrix inverse in the algorithm, possible implementations on massively parallel machines are proposed. Even the huge number of processors available on massively parallel machines cannot make the inversion problem tractable without additional assumptions. For practical radar imaging problems, the matrix to be inverted is Toeplitz-Block Toeplitz, so some savings in computations are possible by exploiting this structure. Other improvements are possible by making further assumptions about the transmitted signal and the image to be formed. Finally simulations are presented and the need for regularization methods discussed.

A model for the received signal in a radar system is given in [5] for reflections at microwave frequencies and in [6] for reflections at optical frequencies. The reflectivity process which characterizes the

† This work was supported by contract number N00014-89-K-1508 from the Office of Naval Research.
To appear in Proceedings of the 1990 Conference on Information Sciences and Systems, Princeton NJ, March 1990.

target is a random process which is stationary at each delay. Thus the samples of the reflectivity process have Toeplitz covariances, and the problem of estimating the parameters of the underlying spectra reduces to a Toeplitz covariance estimation problem. We take circulant extensions of the covariance matrices and estimate spectrum samples.

1. Statistical Model

The target is described in terms of its reflectivity. It is assumed that the radar transmitted signal is a plane wave at the target so points on the target at a cross-section perpendicular to the line of sight sum up to contribute to the same return signal. This sum changes as a function of time and is denoted by $b(t,\tau)$. The variable τ is the distance of points on the target from the transmitter given in time units as the time it takes for a wave to propagate to the target and back to the transmitter (two-way delay). From this model, it is not apparent that separate points at the same two-way delay τ may be differentiated to obtain an image. These points may be differentiated if their velocities relative to the transmitter are different. In particular, if the target is a rotating rigid body, then the velocity of a point in the direction of the transmitter is proportional to the distance of that point from the line of sight. Since the Doppler shift introduced by a point on the target is proportional to this directed velocity, a delay-Doppler image of a rotating rigid body is equivalent to a range-crossrange image. The problem is to determine the power reflected from points as a function of delay and Doppler and then to display this power function as an image of the target. Even if this rigid body assumption for the target is not valid, a delay-Doppler image can be a useful image of the target region.

The reflectivity $b(t,\tau)$ is the superposition of all reflectivities at delay τ, times the Doppler shift terms they introduce. It may be expressed as

$$b(t,\tau) = \int_{-\infty}^{\infty} c(f,\tau) e^{j2\pi ft}\, df\,, \tag{1.1}$$

where $c(f,\tau)$ is the reflectivity of the points on the target at two-way delay τ which introduce a Doppler shift f. The target is assumed to have finite extent. This implies that both $b(t,\tau)$ and $c(f,\tau)$ are zero for τ outside of some fixed interval. It also implies that $c(f,\tau)$ is zero for f outside of some finite interval because the Doppler variable corresponds to crossrange extent of the target.

The literature on radar reflections [5,6] describes statistical models for the reflectivity when the target is rough in the sense that multiple scattering sites are present in a resolution patch on the target. The model states that the reflectivity of a patch on the target is a complex valued Gaussian random variable with zero mean and is uncorrelated from patch to patch. When our model is discretized with $I_{CR} \times I_R$ resolution cells (the subscripts CR and R denote *cross-range* and *range*, respectively), this corresponds to the assumptions on the reflectivity of patches of the target

$$\begin{aligned} E[c(k,i)\,c(k',i')] &= 0 \\ E[c(k,i)\,c^*(k',i')] &= \sigma(k,i)\,\delta_{kk'}\,\delta_{ii'}\,, \end{aligned} \tag{1.2}$$

for $-(I_{CR}-1)/2 \le k \le (I_{CR}-1)/2$, $0 \le i \le I_R - 1$. Here, $\sigma(k,i)$ is the real, nonnegative covariance of the reflectivity of the patch at Doppler k and delay i. The scatterers described by this model are called wide-sense stationary uncorrelated scatterers (WSSUS) by Van Trees [5] because the assumptions imply that

$$E[b(n,i)b^*(n',i')] = K_b(n-n',i)\,\delta_{ii'}\,, \tag{1.3}$$

where $b(n,i) = b(n\,\Delta t - i\,\Delta\tau/2, \Delta\tau)$ and

$$K_b(n,i) = \sum_{k=-(I_{CR}-1)/2}^{(I_{CR}-1)/2} \sigma(k,i) \exp[j\,2\pi k n f_D\,\Delta t]\,, \tag{1.4}$$

where f_D is the step between successive Doppler samples in the image plane, $\Delta\tau$ is the delay step, and Δt is the time between successive samples of $r(t)$.

In a radar system, each signal is typically described as the product of a baseband signal times a complex exponential at the carrier frequency. All of the interactions of interest for narrow band radar systems

may be described in terms of these complex valued baseband signals which will be called complex envelopes of the signals or simply the signals.

Let $s_T(t)$ and $s_R(t)$ be the complex envelopes of the transmitted and received signals, respectively. The received signal is the superposition of the reflectivity from all two-way delays times the appropriate transmitted signals,

$$s_R(t) = \int_{-\infty}^{\infty} s_T(t-\tau) b(t-\tau/2,\tau) d\tau. \tag{1.5}$$

In Equation (1.5), the $\tau/2$ arises because it takes that long for the reflected signal to return to the receiver. The available data are the sum of the radar return signal and additive white Gaussian noise

$$r(t) = s_R(t) + w(t)$$

$$r(t) = \int_{-\infty}^{\infty} s_T(t-\tau) b(t-\tau/2,\tau) d\tau + w(t) \ . \tag{1.6}$$

This equation forms the basis for our model.

The processing is assumed to be performed digitally so that discretized versions of b and c are used. Equation (1.1) is substituted into (1.6) and approximated by samples of the signals c and w. The discrete equation may be written in vector form as

$$\mathbf{r} = \Gamma^t \mathbf{c} + \mathbf{w}, \tag{1.7}$$

where Γ is a matrix whose entries are samples of the transmitted signal times appropriate complex exponentials, and the vectors $\mathbf{c}$ and $\mathbf{w}$ are defined appropriately.

2. Maximum-Likelihood Solution

The loglikelihood function for the data is

$$L(\mathbf{K}_R;\mathbf{r}) = -\ln(\det \mathbf{K}_R) - \mathbf{r}^t (\mathbf{K}_R)^{-1} \mathbf{r}, \tag{2.1}$$

where $\mathbf{K}_R$ is the covariance matrix for the received data $\mathbf{r}$,

$$\mathbf{K}_R = \Gamma^t \Sigma \Gamma + N_0 \mathbf{I}, \tag{2.2}$$

and Σ is a matrix with diagonal entires $\sigma(i,k)$. As shown first by Burg, et al. [3], a necessary condition for a matrix to maximize the loglikelihood function is the trace condition

$$tr\,[(\mathbf{K}_R^{-1}\mathbf{r}\mathbf{r}^t\mathbf{K}_R^{-1} - \mathbf{K}_R^{-1})\delta\mathbf{K}_R] = 0. \tag{2.3}$$

The matrix $\delta\mathbf{K}_R$ is a variational matrix which takes values in all possible additive variations of the matrix $\mathbf{K}_R$, and may be rewritten as $\Gamma^t\delta\Sigma\Gamma$, where $\delta\Sigma$ is a diagonal variational matrix.

An EM algorithm has been derived [4] to estimate the matrix Σ. The estimate at step p+1 of Σ is given by the diagonal elements of the conditional expectation of $\mathbf{c}\mathbf{c}^t$ or

$$\Sigma^{(p+1)} \overset{d}{=} E\,[\mathbf{c}\mathbf{c}^t | \Sigma^{(p)}, \mathbf{r}], \tag{2.4}$$

where $\overset{d}{=}$ means that the off diagonal terms on the left are zero and the diagonal terms on the left equal the diagonal terms on the right. The computation indicated in (2.4) is a standard problem in estimation theory. This equation may be written as [4]

$$\Sigma^{(p+1)} \overset{d}{=} \Sigma^{(p)} + \tag{2.5}$$

$$\Sigma^{(p)}\Gamma(\Gamma^t\Sigma^{(p)}\Gamma + N_0\mathbf{I})^{-1}(\mathbf{r}\mathbf{r}^t - \Gamma^t\Sigma^{(p)}\Gamma - N_0\mathbf{I})(\Gamma^t\Sigma^{(p)}\Gamma + N_0\mathbf{I})^{-1}\Gamma^t\Sigma^{(p)}.$$

Equation (2.5) defines the iteration sequence used by the computations. To start the algorithm, $\Sigma^{(0)}$ is chosen as an arbitrary positive definite diagonal matrix.

There are two possible images which may be displayed. The first consists of estimates of samples of the scattering function. The second consists of magnitudes or squared magnitudes of estimates of samples of the reflectivity function. The diagonal elements of Σ are the values which are displayed as the scattering function image of the target. Thus, at each stage an image is calculated and may be displayed. Some appropriate stopping criterion is used to terminate the algorithm. At each stage of the algorithm the conditional mean estimate of the reflectance is also generated. At step p, this estimate is

$$E[\mathbf{c}|\Sigma^{(p)},\mathbf{r}] = \Sigma^{(p)}\Gamma(\Gamma^{\dagger}\Sigma^{(p)}\Gamma + N_0\mathbf{I})^{-1}\mathbf{r}. \tag{2.6}$$

The magnitude or the magnitude squared of c may also be displayed at each stage of the algorithm as an image of the target. Thus both types of radar image commonly viewed are generated by our algorithm. We feel this is a unique feature of our algorithm.

Let the corresponding sequence of covariance matrices for the data $\mathbf{r}$ be denoted $\mathbf{K}_R^{(p)} = \Gamma^{\dagger}\Sigma^{(p)}\Gamma + N_0\mathbf{I}$. Since this iteration is an EM algorithm, it has all of the properties associated with this type of algorithm. In particular, the incomplete data loglikelihood is nondecreasing in the sequence of covariance matrices $\mathbf{K}_R^{(p)}$.

There are obviously issues associated with the appropriate or desired sampling rates of $r(t)$ and of $c(f,\tau)$. Some of these issues are addressed in [4]. The quality of the image obtained and the resolution achievable are intimately related to the sampling issues.

This section has presented a review of the equations used to produce target images. This approach starts with a model which accurately accounts for the random nature of radar reflections and adopts the maximum likelihood method of statistics to estimate delay-Doppler high resolution images of radar targets. Questions associated with uniqueness of spectrum estimates and convergence of the EM algorithm are discussed in [4,7]. Equation (2.5) is a computationally demanding update. Some of the issues associated with this update are addressed in the next section.

3. Implementation of Algorithm

The implementation of the iterative algorithm described in (2.5) involves a number of linear algebra operations on arrays which may be very large. As a result, the development of efficient and numerically stable routines, on both serial and massively-parallel machines, to perform these operations is of great importance in evaluating this algorithm. Because massively-parallel architectures are relatively new and higher level software is not yet available, algorithmic development for these machines is more involved.

In order to understand the issues associated with our effort to apply the algorithm of (2.5) to a parallel architecture, one must understand, to a degree, the limitations imposed by our resident parallel machine, the DAP 510, and its programming language, Fortran-Plus. Fortran-Plus is an adapted version of Fortran which hides all communication between processor elements from the user by virtue of its data structures. Scalars, length-32 vectors, and 32×32 matrices are all distinct data types, and arrays of each may be formed. Additionally, Fortran-Plus supports no complex data type. Thus, the large, complex matrices found in our signal model should be represented as 3-dimensional arrays of Fortran-Plus matrices in order to take maximum advantage of the parallel nature of the DAP. As we shall see, this in some ways complicates and in other ways simplifies the task of programming.

3.1 Algebraic Operations in Algorithm

If one examines (2.5), it is seen that the indicated iterations can be performed by making use of only the following basic linear algebra operations:

$$\mathbf{C} = \mathbf{A} + \mathbf{B}$$

$$\mathbf{C} = \mathbf{AB}$$

$$\mathbf{C} = \mathbf{A}^{\dagger}\mathbf{B}$$

$$\mathbf{A} = \mathbf{v}\mathbf{w}^{\dagger}$$

$$\mathbf{w} = \mathbf{A}^{\dagger}\mathbf{v}$$

$$\mathbf{B} = \mathbf{A}^{-1} \tag{3.1}$$

where **A**, **B**, and **C** are complex matrices, and **v** and **w** are complex vectors. The addition operation is greatly simplified by the data structures of Fortran-Plus; one need only add each of the 32×32 components of **A** to the corresponding component of **B**. Furthermore, each of these 32×32 additions can be accomplished by a single statement in Fortran-Plus.

Each of the four multiplication operations, although different, can be performed in Fortran-Plus by following a similar strategy. Consider the first of these, that of performing matrix multiplication, where the matrices in question are complex and of size $32n \times 32n$, where n is a integer greater than 1. Thus, these matrices would be represented as an $n \times n \times 2$ array of Fortran-Plus matrices.

Golub and Van Loan [8] discuss two approaches to the matrix multiplication problem based on the following partitioning:

$$n \times n \times 2 \tag{3.2}$$

The more efficient of these two methods is that developed by Strassen, which performs the multiplication of two $M \times M$ matrices via 7 multiplies and 18 adds of $M/2 \times M/2$ matrices:

$$\begin{aligned}
\mathbf{P}_1 &= (\mathbf{A}_{11} + \mathbf{A}_{22})(\mathbf{B}_{11} + \mathbf{B}_{22}) \\
\mathbf{P}_2 &= (\mathbf{A}_{21} + \mathbf{A}_{22})\mathbf{B}_{11} \\
\mathbf{P}_3 &= \mathbf{A}_{11}(\mathbf{B}_{12} - \mathbf{B}_{22}) \\
\mathbf{P}_4 &= \mathbf{A}_{22}(\mathbf{B}_{21} - \mathbf{B}_{11}) \\
\mathbf{P}_5 &= (\mathbf{A}_{11} + \mathbf{A}_{12})\mathbf{B}_{22} \\
\mathbf{P}_6 &= (\mathbf{A}_{21} - \mathbf{A}_{11})(\mathbf{B}_{11} + \mathbf{B}_{12}) \\
\mathbf{P}_7 &= (\mathbf{A}_{12} - \mathbf{A}_{22})(\mathbf{B}_{21} + \mathbf{B}_{22}) \\
\mathbf{C}_{11} &= \mathbf{P}_1 + \mathbf{P}_4 - \mathbf{P}_5 + \mathbf{P}_7 \\
\mathbf{C}_{12} &= \mathbf{P}_3 + \mathbf{P}_5 \\
\mathbf{C}_{21} &= \mathbf{P}_2 + \mathbf{P}_4 \\
\mathbf{C}_{22} &= \mathbf{P}_1 + \mathbf{P}_3 - \mathbf{P}_2 + \mathbf{P}_6
\end{aligned} \tag{3.3}$$

This method is particularly useful for application to the DAP, if used recursively to reduce the problem to that of multiplying and adding a series of 32×32 real matrices, which are straightforward and efficient operations in Fortran-Plus. A slight modification of Strassen's algorithm is required to handle the separate real and imaginary submatrices, but the recursive partitioning scheme remains intact.

The issue of matrix inversion within the algorithm is more fundamental, as the choice of algorithm is not obvious. Press, et. al. [16] discuss the merits of Strassen inversion, which is based on an analogous partitioning.

$$\begin{aligned}
&32 \times 32 \\
\mathbf{R}_1 &= \mathbf{A}_{11}^{-1} \\
\mathbf{R}_2 &= \mathbf{A}_{21} \times \mathbf{R}_1 \\
\mathbf{R}_3 &= \mathbf{R}_1 \times \mathbf{A}_{12} \\
\mathbf{R}_4 &= \mathbf{A}_{21} \times \mathbf{R}_3 \\
\mathbf{R}_5 &= \mathbf{R}_4 - \mathbf{A}_{22} \\
\mathbf{R}_6 &= \mathbf{R}_5^{-1} \\
\mathbf{C}_{12} &= \mathbf{R}_3 \times \mathbf{R}_6 \\
\mathbf{C}_{21} &= \mathbf{R}_6 \times \mathbf{R}_2
\end{aligned} \tag{3.4}$$

$$\mathbf{R}_7 = \mathbf{R}_3 \times \mathbf{C}_{21}$$
$$\mathbf{C}_{22} = -\mathbf{R}_6 \tag{3.5}$$

The recursive partitioning associated with this algorithm makes it particularly attractive for implementation on the DAP, as all numerical operations can be performed on Fortran-Plus matrices.

3.2 Further matrix inversion issues

In addition to the estimated covariance matrix $\mathbf{K}_R$, often the matrix Γ is inverted for the purpose of comparing the results of the EM algorithm with those obtained from (1.7) in an attempt to reconstruct c using linear processing. This is in the spirit of conventional processing which is based on linear operations. In addition, this corresponds to the maximum likelihood solution shown in (4.6) in the special case discussed there. Note that the parallel algorithm for inverting matrices in (3.5) seeks to invert the upper left block of the original matrix. It happens that, under certain conditions, the upper left block of the matrix Γ may be singular, or at least poorly conditioned. As the partitioning is repeated, the algorithm becomes numerically unstable for a large class of matrices Γ. Other algorithms which first partition the matrix to be inverted into its real and imaginary parts suffer the same fate.

If one examines the literature, one finds a wide variety of inversion algorithms based on decompositions. Each seeks to make the problem more tractable by factoring the matrix in question in such a way that the factors are easily inverted. Many of these decompositions require that the matrix to be inverted have a special structure which either $\mathbf{K}_R$ or Γ do not satisfy. For example, the LU factorization requires that all principal submatrices of the matrix in question be non-singular, in order for the algorithm to be stable [8]. As we have already seen, this may not be true for Γ.

We have chosen to implement the QR decomposition in Fortran-Plus for this purpose. This routine factors a square matrix **A** in in the following way:

$$\mathbf{A} = \mathbf{QR} \tag{3.6}$$

such that **Q** is unitary ($\mathbf{Q}^\dagger\mathbf{Q} = \mathbf{I}$) and **R** is upper triangular. This gives rise to the following inversion algorithm:

$$\mathbf{A}^{-1} = \mathbf{R}^{-1}\mathbf{Q}^\dagger \tag{3.7}$$

Thus the inversion problem has been transformed to the separate problems of calculating the factors, inverting a triangular matrix, and performing one matrix multiplication. Note that although this process does not reduce the complexity of the inversion problem, it defines the solution in terms of operations which are well known.

Unfortunately, the parallel implementation of this inversion strategy is not as computationally efficient as one might hope, as the operations associated with the QR decomposition can really only be applied to columns of data at a time. In order to take maximum advantage of the DAP's parallel architecture, such operations should be applied to 32×32 matrices at a time. However, the inverse of Γ is computed only once for the purposes of forming the output of conventional processing, while the inverse of $\mathbf{K}_R$ is computed at every iteration of the EM algorithm. Therefore, a significant increase in the speed of the algorithm as a whole may be realized if a more efficient parallel routine for computing $\mathbf{K}_R$ is utilized. Certain special cases for the transmitted signal give rise to such situations, as discussed in the next section.

4. Improvements in Complexity

From the considerations developed in Section 3, it appears that the task of inverting the covariance matrix $\mathbf{K}_R$ is a formidable one. In this section we examine under what conditions the complexity of this operation is lower than first expected.

4.1 Toeplitz-Block Toeplitz structure

Consider the often-used stepped-frequency waveform [9]. The transmitted signal is made of N_b bursts of N_p pulses each. The signal transmitted in pulse p_n of burst b_n is a complex sinusoid at frequency

$p_n f_T$ (where f_T is the frequency step between successive pulses):

$$s_T(t) = \exp[-j2\pi p_n f_T t] , \tag{4.1}$$

Usually the return signal is sampled at the pulse repetition frequency (1 sample per pulse ; $N = N_b N_p$). If $f_T \Delta t$ is an integer, we obtain

$$s_T(n\Delta t - i\Delta\tau) = \exp[j2\pi p_n f_T i\Delta\tau] , \tag{4.2}$$

with $n = N_p b_n + p_n$, $0 \le b_n < N_b$, $0 \le p_n < N_p$. In this instance $\mathbf{K}_R$ has a special structure, known as Toeplitz-Block-Toeplitz (TBT) [10]. From (2.2):

$$\begin{aligned}
\mathbf{K}_R(n,m) &= \sum_{k=0}^{l_{CR}-1}\sum_{i=0}^{l_R-1} \sigma(k,i)\exp[j2\pi k f_D(n-m)\Delta t]\, s_T(n\Delta t - i\Delta\tau)\, s_T^*(m\Delta t - i\Delta\tau) \\
&= \sum_{k=0}^{l_{CR}-1}\sum_{i=0}^{l_R-1} \sigma(k,i)\exp[j2\pi k f_D(n-m)\Delta t]\exp[j2\pi(p_n-p_m) f_T i\Delta\tau] \\
&= \sum_{k=0}^{l_{CR}-1}\sum_{i=0}^{l_R-1} \sigma(k,i)\exp[j2\pi k f_D\,(N_p(b_n-b_m)+(p_n-p_m))\Delta t]\times \\
&\qquad \exp[j2\pi(p_n-p_m) f_T i\Delta\tau] .
\end{aligned} \tag{4.3}$$

Thus $\mathbf{K}_R(n,m)$ is a function of $b_n - b_m$ and $p_n - p_m$ only. $\mathbf{K}_R$ possesses a doubly Toeplitz structure: $\mathbf{K}_R$ is made of $N_b \times N_b$ Toeplitz blocks of dimension $N_p \times N_p$, themselves arranged in a Toeplitz structure.

Special algorithms have been developed for inverting TBT matrices [10,11]. The complexity of Wax's algorithm is min $(N_p^2 N_b^3, N_p^3 N_b^2)$. For $N_b = N_p$ this is equal to $N^{5/2}$, which is significantly smaller than N^3 for large N. The complexity of the TBT matrix inversion problem on a parallel machine has not yet been investigated.

4.2 Special choice of the parameters

In the special case where $I_R I_{CR} = N$ and $\Gamma\Gamma^\dagger$ is the identity matrix, the covariance matrix (2.2) has the form

$$\mathbf{K}_R = \Gamma^\dagger(\Sigma + N_0\mathbf{I})\Gamma . \tag{4.4}$$

The trace condition (2.3) becomes

$$Tr\,[\delta\Sigma(\Sigma+N_0\mathbf{I})^{-1}(\mathbf{y}\mathbf{y}^\dagger - \Sigma - N_0\mathbf{I})(\Sigma+N_0\mathbf{I})^{-1}] = 0 , \tag{4.5}$$

where

$$\mathbf{y} = (\Gamma^{-1})^\dagger \mathbf{r} . \tag{4.6}$$

The covariance to be inverted in this equation is diagonal. Numerically the inversion problem is now trivial. Actually this equation can even be solved in closed form. The maximum-likelihood estimates are given by:

$$\hat{\sigma}(k,i) = \max\left[|y(k,i)|^2 - N_0 , 0\right] . \tag{4.7}$$

It should be noted that for stepped-frequency waveforms, the design $I_{CR} = N_b$ and $I_R = N_p$ leads to $\Gamma\Gamma^\dagger$ being the identity matrix. In this instance, $|y|^2$ is the output of the conventional processing for the radar return, based on Fourier transforms [9]. The ML estimator is obtained by subtracting N_0 from this estimate, followed by setting negative values to zero.

5. Regularization Methods

The ML estimator presented in Section 2 has been shown to offer lower bias and better resolution than conventional estimators. However the estimates obtained with the estimation setup of Section 2 are not mean-square consistent. This unfortunate result is due to the the number of parameters ($I_{CR} I_R$) being at least equal to the number of data (N) in the method used. We are actually dealing with a small-sample

MLE problem. The situation becomes worse as $I_R I_{CR}$ is increased in an attempt to improve resolution. The estimates are extremely sensitive to small changes in the data and exhibit a very rough shape characterized by very sharp peaks and low valleys.

Such a behavior is of course highly undesirable, and we need a *regularization* of the estimates. This concept was introduced by Tikhonov [12].

One possible approach for regularization of the estimates is the method of *sieves* introduced by Grenander [13, p. 357]. A sieve in a parameter space **A** is a family of subsets $S(\mu)$ of **A** indexed by a positive parameter μ called the *mesh size*. A restricted ML estimate exists over each set $S(\mu)$. As the mesh size tends to zero the sets $S(\mu)$ will be large enough to allow the ML solution to converge to any solution in **A**. For the problem at hand the choice of a "good" value of μ will depend on the data record size and the noise level, among other factors.

A major problem is to find "good" sieves, which will make the ML estimates converge and possess some desirable practical features such as analytical and/or computational tractability. Recently we proposed a method of sieves for a spectrum estimation problem [14]. The function to be estimated is represented by a series expansion, and the restricted set $S(Q)$ is the set of series truncated to Q terms. Essentially the variational matrix δK_R is now constrained in a Q-dimensional subset which will be much smaller than the set of all possible variations. This method can be extended to the radar imaging problem as follows [15].

5.1 Definition of a Sieve

The scattering function $\sigma(k,i)$ can be viewed as a vector in the Euclidean space $\mathbf{R}^{I_{CR} I_R}$. Let $\psi_{lm}(k,i)$, $0 \le l < I_{CR}$, $0 \le m < I_R$, be a set of $I_{CR} I_R$ positive basis vectors in $\mathbf{R}^{I_{CR} I_R}$. It is not required that this set be orthogonal. Consider the set of all vectors in the span of $\{\psi_{lm}\}$ with nonnegative coefficients:

$$\mathbf{A} := \left\{ \sigma(k,i) = \sum_{l=0}^{I_{CR}-1} \sum_{m=0}^{I_R-1} a(l,m)\,\psi_{lm}(k,i) \;,\; a(l,m) \ge 0 \right\}. \tag{5.1}$$

We define our parameter set to be **A**. The vectors in this set are represented by means of the coefficients $a(l,m)$ $0 \le l < I_{CR}$, $0 \le m < I_R$. We define a sieve $S(Q_{CR}, Q_R)$ on $\sigma(k,i)$ by truncating the series representation (5.1) to $Q_{CR} Q_R$ terms. The integers Q_{CR} and Q_R are constrained to be no larger than I_{CR} and I_R, respectively.

$$S(Q_{CR}, Q_R) := \left\{ \sigma(k,i) = \sum_{l=0}^{Q_{CR}-1} \sum_{m=0}^{Q_R-1} a(l,m)\,\psi_{lm}(k,i) \,,\, a(l,m) \ge 0 \right\}. \tag{5.2}$$

The vectors in this subset are represented by means of $Q_{CR} Q_R$ coefficients $a(l,m)$ $0 \le l < Q_{CR}$, $0 \le m < Q_R$. We view these coefficients as being $Q_{CR} Q_R$ unknown parameters for which ML estimates are sought. In [15] we show that consistency of the estimates can be obtained if Q_{CR} and Q_R grow at an appropriate rate with N. Typically the basis functions that are designed are *localized* in range and cross-range (i.e. k and i). Next we show how the ML estimates of the coefficients are computed.

5.2 Algorithm No.1

For each l, m, define the support set $\mathbf{D}_{lm}$ of $\psi_{lm}(k,i)$. Also construct a diagonal matrix Ψ_{lm} made of the samples $\psi_{lm}(k,i)$ and a *basis covariance matrix*

$$\mathbf{K}_{lm} := \Gamma^{\dagger} \Psi_{lm} \Gamma \,. \tag{5.3}$$

It follows that $\mathbf{K}_R$, subject to the sieve constraint, assumes the form

$$\mathbf{K}_R = \sum_{l=0}^{Q_{CR}-1} \sum_{m=0}^{Q_R-1} a(l,m)\,\mathbf{K}_{lm} + N_0 \mathbf{I} \,. \tag{5.4}$$

The estimation problem is now to maximize the loglikelihood (2.1) subject to the constraint (5.4). From

(2.3) and (5.4), a necessary condition for $\mathbf{K}_R$ to maximize (2.1) is the trace condition

$$tr\,[\,(\mathbf{K}_R^{-1}\mathbf{r}\mathbf{r}^{\dagger}\mathbf{K}_R^{-1}-\mathbf{K}_R^{-1})\,\mathbf{K}_{lm}\,]=0\,,\quad 0\le l<Q_{CR}\,,\ 0\le m<Q_R\,. \tag{5.5}$$

The maximization problem is solved using an EM algorithm similar to the one derived in [14]. The complete data are defined as $(\{c_{lm}\},\ \mathrm{w}\ )$, where c_{lm} , $0\le l<Q_{CR}$, $0\le m<Q_R$, is an $I_{CR}I_R$-vector with covariance $a(l,m)\,\Psi_{lm}$ known up to the scale factor $a(l,m)$. The estimate at step $p+1$ of the coefficient $a(l,m)$ is given by

$$\hat{a}(l,m)^{(p+1)}=\frac{1}{|\mathbf{D}_{lm}|}\sum_{k,i\in\mathbf{D}_{lm}}\frac{E\,[|c_{lm}(k,i)|^2|r,\hat{a}(l,m)^{(p)}]}{\psi_{lm}(k,i)}\,. \tag{5.6}$$

Evaluating the expectation yields the update equation

$$\hat{a}(l,m)^{(p+1)}=\hat{a}(l,m)^{(p)}+ \tag{5.7}$$
$$\frac{1}{|\mathbf{D}_{lm}|}\left[\,\hat{a}(l,m)^{(p)}\right]^2 tr\,[\,(\mathbf{K}_R^{(p)-1}\mathbf{r}\mathbf{r}^{\dagger}\mathbf{K}_R^{(p)-1}-\mathbf{K}_R^{-1})\,\mathbf{K}_{lm}\,]\,.$$

The algorithm is started with positive coefficients $a(l,m)$. Each matrix $\mathbf{K}_{lm}$ is constant and can be computed off-line. At each step the covariance matrix $\mathbf{K}_R^{(p)}$ must be inverted, which accounts for N^3 operations. Following this, updating each $\hat{a}(l,m)$ just requires N^2 multiplications/additions. The complexity of each step of the EM algorithm is then

$$N^3+Q_{CR}Q_RN^2\,, \tag{5.8}$$

as compared to $N^3+I_{CR}I_RN^2$ in the unconstrained case of §2. This saving is important, as it is now possible to let $I_{CR}I_R$ tend to infinity. However the problem of inverting the covariance matrix is just as difficult as before. In the next section we show how the method is handled in a special case of interest.

5.3 Algorithm No.2

The regularization method described in §5.2 can be applied to the special case presented in §4.2. In general the trace condition cannot be solved in closed form, yet the computational requirements are drastically reduced. The update equation (5.7) becomes

$$\hat{a}(l,m)^{(p+1)}=\hat{a}(l,m)^{(p)}+\frac{1}{|\mathbf{D}_{lm}|}\left[\,\hat{a}(l,m)^{(p)}\right]^2\times$$
$$tr\,[\,((\Sigma^{(p)}+N_0\mathbf{I})^{-1}\mathbf{y}\mathbf{y}^{\dagger}(\Sigma^{(p)}+N_0\mathbf{I})^{-1}-(\Sigma^{(p)}+N_0\mathbf{I})^{-1})\,\Psi_{lm}\,]$$
$$=\hat{a}(l,m)^{(p)}+\frac{1}{|\mathbf{D}_{lm}|}\left[\,\hat{a}(l,m)^{(p)}\right]^2\times \tag{5.9}$$
$$\sum_{k,i\in\mathbf{D}_{lm}}\frac{|\mathbf{y}(k,i)|^2-\sigma(k,i)^{(p)}-N_0}{(\sigma(k,i)^{(p)}+N_0)^2}\,\psi_{lm}(k,i)\,.$$

where the vector $\mathbf{y}$ is defined in (4.6). The matrix $\mathbf{K}_R$ has been reduced to a diagonal form and the numerical complexity of each iteration is equal to

$$\sum_{l=0}^{Q_{CR}-1}\sum_{m=0}^{Q_R-1}|\mathbf{D}_{lm}|$$

multiplications/additions. This number is the total area of the support sets of the basis functions. For a discussion on the design of the basis functions and their support sets we refer to [15]. Typically it is required that the support sets cover the entire range and cross-range domain, so that the complexity is equal to

$$\sum_{l=0}^{Q_{CR}-1}\sum_{m=0}^{Q_R-1}|\mathbf{D}_{lm}|\;=\;\lambda I_{CR}I_R\;=\;\lambda N\,, \tag{5.10}$$

where the constant $\lambda\ge 1$ indicates the amount of overlapping of the support sets. This figure is a major improvement from the corresponding figure in (5.8).

The special case that was treated in this section is very important, for it can be shown that the estimates obtained with the algorithms of Sections 5.2 and 5.3 using different values for $I_{CR}I_R$ are asymptotically equivalent as $N \rightarrow \infty$ [15]. This result justifies the choice of setting $I_{CR}I_R = N$, followed by using the latter algorithm, as a valid approach to estimating $\sigma(k,i)$.

6. Simulations

We have performed simulations which will allow us to compare the results of the algorithm described in section 2, the algorithm of section 5, as well as the results of conventional processing. In the first of these, we designed a series of routines for a massively parallel machine which would generate simulated radar data according to the indicated diffuse target model, and then implement the algorithms of section 2 to estimate the target image. Our results here were quite promising. In every case examined, our algorithm outperformed the conventional processing, and was especially superior in those cases where the noise level was significant.

There are two sets of plots shown below. They are computed for a scattering function which is a square of four by four pixels centered in a square of size eight by eight pixels. The intensities of the scattering function in each of the regions and of the noise are the parameters in the plots shown. The transmitted signal is a stepped frequency waveform, with one sample taken per pulse. A total of 64 pulses are transmitted in the simulations with eight bursts of eight pulses each. The frequency step is chosen such that $f_T \Delta\tau$ equals 0.125.

The first plots show the convergence of the value of the loglikelihood function for single realizations of the simulation at two background noise levels. The signal to noise ratio indicated on the plots is computed by first taking the ratio of the expected total energy in the signal part of the received signal and dividing by the expected total energy in the noise, and then computing 10 times the log of that number.

The second set of plots shows the summed absolute error between the maximum likelihood image assuming all of the data (c) is observed and the maximum likelihood image assuming $\mathbf{r}$ is observed. These plots measure the degradation in performance due to the noise and having to invert Γ. These plots show that the actual values displayed in the image continue to change even after the likelihood function is close to its maximum. Visually, this effect is manifested in the image becoming rougher as the iterations proceed. The regularization methods discussed in section 5 and in [15] take into account more prior information known about the scattering function and thus produce less rough images.

References

[1] A.D. Dempster, N.M. Laird, D.B. Rubin, "Maximum-Likelihood from Incomplete Data via the EM Algorithm", *J. Royal Stat. Soc., vol. B39, pp. 1-37, 1977.*

[2] M. I. Miller, D. L. Snyder, "The Role of Likelihood and Entropy in Incomplete-Data Problems: Applications to Estimating Point-Process Intensities and Toeplitz Constrained Covariances," *Proc. IEEE, vol.75*, No.7, July 1987.

[3] J.P. Burg, D.G. Luenberger, D.L. Wenger, "Estimation of Structured Covariance Matrices", *Proc. IEEE, vol. 70*, pp. 963-974, Sept. 1982.

[4] D. L. Snyder, J. A. O'Sullivan, M. I. Miller, "The Use Of Maximum-Likelihood Estimation For Forming Images Of Diffuse Radar-Targets From Delay-Doppler Data," *IEEE Trans. on Information Theory, vol. 35*, No. 3, May 1989, pp.536-548.

[5] H.L. Van Trees, "Detection, Estimation, and Modulation Theory, Part I", *Wiley and Sons, New York 1968.*

[6] J. Shapiro, B.A. Capron, and R. C. Harney, "Imaging and Target Detection with a Heterodyne-Reception Optical Radar," *Applied Optics*, vol. 20, pp. 3293-3313, 1981.

[7] J. A. O'Sullivan, P. Moulin, D. L. Snyder, "Cramer-Rao Bounds for Constrained Spectrum Estimation with Application to a Problem in Radar Imaging", *Proc. 26th Allerton Conference on Communication, Control, and Computing*, Urbana-Champaign, IL, 1988, pp. 27-34.

[8] C.H. Golub, C.F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1989.

[9] D.R. Wehner, *High Resolution Radar*, Artech House, Norwood, MA, 1987

[10] M. Wax and T. Kailath, Efficient Inversion of Toeplitz-Block Toeplitz Matrix, *IEEE Trans. on ASSP, vol. 31*, No 5, 1983.

[11] H. Akaike, Block Toeplitz Matrix Inversion, *SIAM J. Applied Math., vol. 24*, No. 2, 1973.

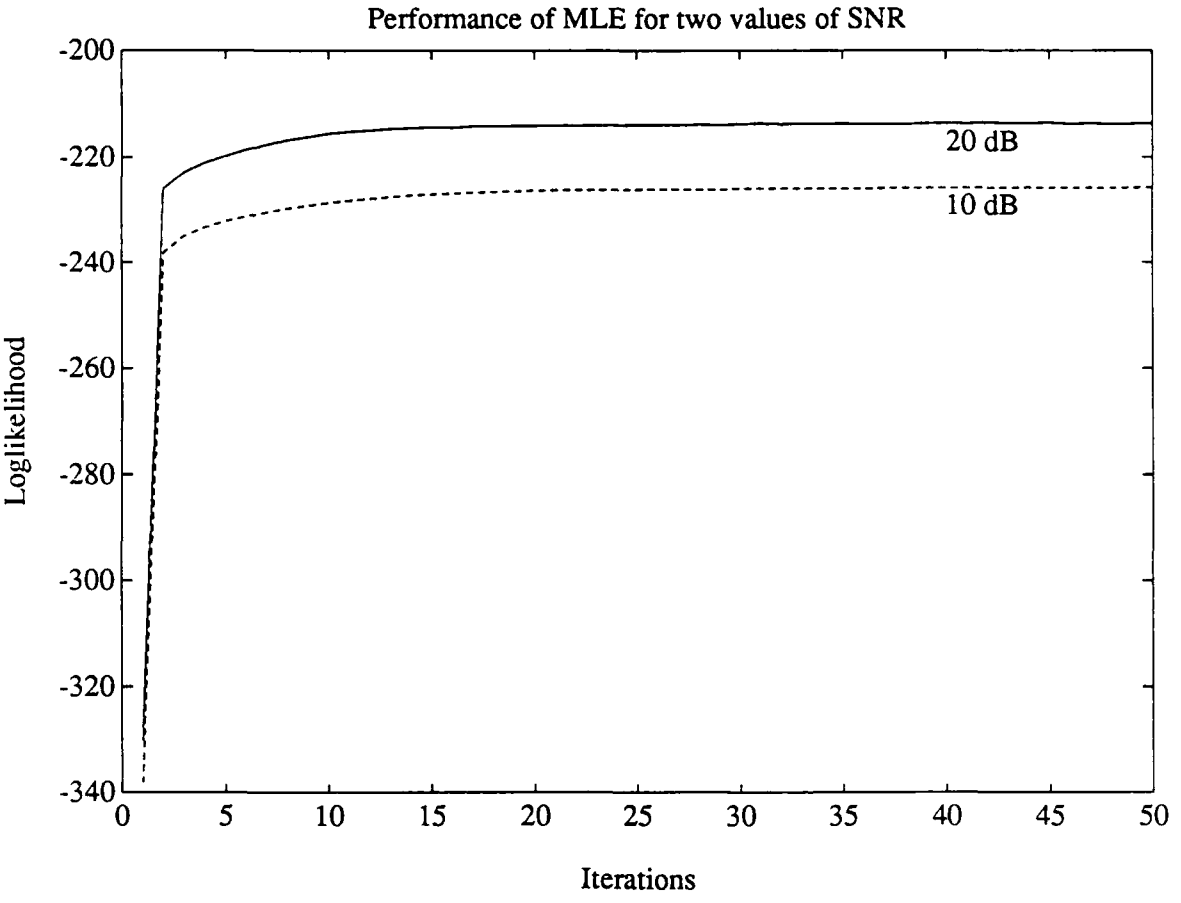
[12] A. Tikhonov, V. Arsenin, *Solutions of Ill-Posed Problems*, Winston & Sons, Washington D.C. 1977.

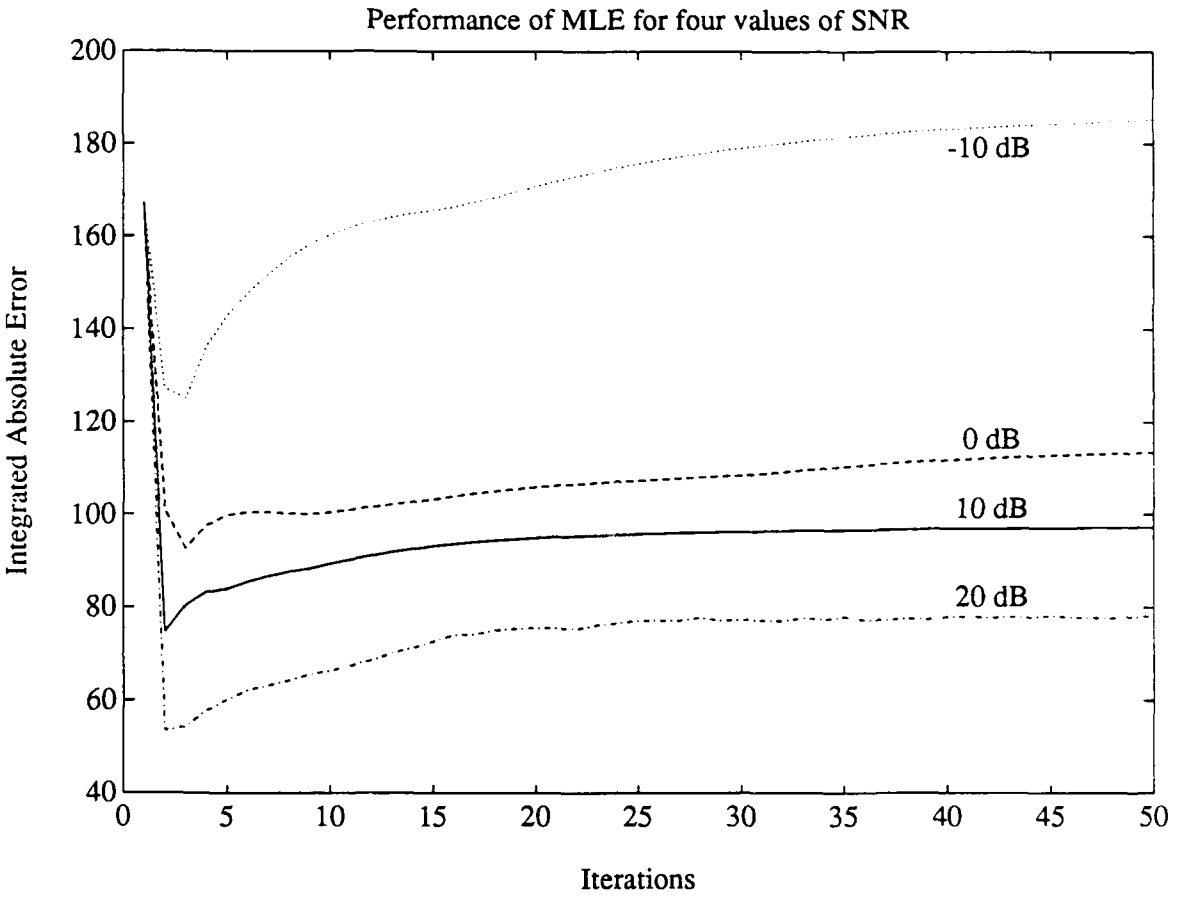
[13] Grenander U., *Abstract Inference*, Wiley 1981.

[14] P. Moulin, D.L. Snyder, J.A. O'Sullivan, "A Sieve-Constrained Maximum-Likelihood Estimator for the Spectrum of a Gaussian Process," *Proc. 27th Allerton Conference on Communication, Control, and Computing*, Urbana-Champaign, IL, 1989.

[15] P. Moulin, *D.Sc. Dissertation*, Washington University, St. Louis, May 1990.

[16] W. H. Press, et al., *Numerical Recipes in C: The Art of Scientific Computation*, Cambridge University Press, Cambridge, 1988.


Performance of MLE for two values of SNR
-200
-220
-240
-260
-280
-300
-320
-340
0
5
10
15
20
25
30
35
40
45
50
20 dB
10 dB
Log likelihood
Iterations


Performance of MLE for four values of SNR
Integrated Absolute Error
Iterations
-10 dB
0 dB
10 dB
20 dB
200
180
160
140
120
100
80
60
40
0
5
10
15
20
25
30
35
40
45
50

6.6 Preprint of: J. A. O'Sullivan and D. L. Snyder, "High Resolution Radar Imaging Using Spectrum Estimation Methods," Proc. August 1989 Program on Signal Processing, Institute for Mathematics and Its Applications, University of Minnesota, Minneapolis MN, to appear.

HIGH RESOLUTION RADAR IMAGING USING SPECTRUM ESTIMATION METHODS

JOSEPH A. O'SULLIVAN
DONALD L. SNYDER †

Abstract. This paper summarizes a new approach to high resolution radar imaging based on modern spectrum estimation techniques. First a statistical model of the radar reflections which properly accounts for the randomness of reflections by targets which are rough on the order of a wavelength of the carrier frequency is introduced. The model for the radar return signal is valid for all transmitted narrowband radar signals. Equations which generate maximum likelihood estimates for the reflectivity power as a function of delay and Doppler coordinates are derived.

Introduction. This paper presents recent research results in high resolution delay-Doppler radar imaging based on statistical models obtained from the underlying physics of spotlight mode radar reflections. We present solution equations which may be used to process radar return signals to form images. These equations bear a strong resemblance to the equations derived in [1] for the generic Toeplitz covariance estimation problem. This is a result of the fact that the reflectivity process which characterizes the target is a random process which is stationary at each delay. Thus the samples of the reflectivity process have Toeplitz covariances and the problem of estimating the parameters of the underlying spectra reduces to a Toeplitz covariance estimation problem. We take circulant extensions of the covariance matrices and estimate spectrum samples. There are several aspects of this problem which differ significantly from the generic problem, but the underlying problem is to estimate Toeplitz covariance matrices.

The model of the reflectivity process is described first. Since the processing is performed digitally, the discrete form of this model is examined in detail. Next, the manner in which the transmitted signal interacts with the reflectivity to form the radar return signal is presented. The imaging problem is reduced to the problem of estimating the spectrum underlying the reflectivity process given samples of the radar return signal. A necessary condition for the maximum likelihood solution is obtained and an EM algorithm approach to solving for the maximum is taken. The results extend those in [1] in three important ways. First, the samples of the process with a Toeplitz covariance are not available. Instead, the stationary process is multiplied by a signal matrix. Second, the model includes an additive white Gaussian noise. Third, the process of interest is a function of two variables. For each value of the delay variable, the process is Toeplitz. Thus the model from [1] is extended to spectra which change as a function of an independent variable.

Reflectivity process. The target is described in terms of its reflectivity. It is assumed that the radar transmitted signal is a plane wave at the target so points on the target at a cross-section perpendicular to the line of sight sum up to contribute to the same return signal. This sum changes as a function of time and is denoted by $y(t,\tau)$. The variable τ is the distance of points on the target from the transmitter given in time units as the time it takes for a wave to propagate to the target and back to the transmitter (two-way delay). From this model it is not apparent that separate points at the same two-way delay τ may be differentiated to obtain an image. These points may be differentiated if their velocities relative to the transmitter are different. In particular, if the target is a rigid body and is rotating about a point along the line of sight, then the velocity of a point in the direction of the transmitter is proportional to the distance of that point from

† Both authors are with the Electronic Systems and Signals Research Laboratory, Department of Electrical Engineering, Washington University, St. Louis, MO 63130. This work was supported by contract number N00014-86-K-0370 from the Office of Naval Research.

the line of sight. Since the Doppler shift introduced by a point on the target is proportional to this directed velocity, a delay-Doppler image of a rotating rigid body is equivalent to a range-crossrange image. The problem is to determine the power reflected from points as a function of delay and Doppler and then to display this power function as an image of the target. Even if this rigid body assumption for the target is not valid, a delay-Doppler image can be a useful image of the target region.

Since the reflectivity $y(t,\tau)$ is the superposition of all reflectivities at delay τ times the Doppler shift terms they introduce, it may be expressed as

$$y(t,\tau) = \int_{-\infty}^{\infty} c(f,\tau) e^{j2\pi f t} df, \tag{1}$$

where $c(f,\tau)$ is the reflectivity of the points on the target at two-way delay τ which introduce a Doppler shift f. The target is assumed to have finite extent. This implies that both $y(t,\tau)$ and $c(f,\tau)$ are zero for τ outside of some fixed interval. It also implies that $c(f,\tau)$ is zero for f outside of some finite interval because the Doppler variable corresponds to crossrange extent of the target. The processing is assumed to be performed digitally so that discretized versions of y and c are used. Suppose that the resolution cells of f and τ are Δf and $\Delta\tau$ respectively. Let there be I_R bins in the delay or range direction (in delay coordinates, the target is of length $I_R \Delta\tau/2$) and let there be I_{CR} bins in the Doppler or cross range direction (so the target is of width $I_{CR}\Delta f$). If samples of the radar return signal are taken every Δt seconds, Equation (1) may be approximated by

$$y(k\Delta t, n\Delta\tau) = \sum_{m=-(I_{CR}-1)/2}^{(I_{CR}-1)/2} c(m,n) e^{j2\pi mk\Delta f \Delta t}, \tag{2}$$

where

$$c(m,n) = c(m\Delta f, n\Delta\tau)\Delta f. \tag{3}$$

The literature on radar reflections [2,3] describes statistical models for the reflectivity when the target is rough on the order of a wavelength of the carrier signal. The model states that the reflectivity of a patch on the target is a complex valued Gaussian random variable with zero mean and is uncorrelated from patch to patch. For our model this corresponds to the assumptions on the reflectivity of patches of the target

$$E[c(i,k)c(m,n)] = 0$$
$$E[c(i,k)c^*(m,n)] = \sigma(i,k)\delta_{i,m}\delta_{k,n}. \tag{4}$$

Here, $\sigma(i,k)$ is the real, nonnegative covariance of the reflectivity of the patch at delay k and Doppler i. The scatterers described by this model are called wide-sense stationary uncorrelated scatterers (WSSUS) by Van Trees [2] because the assumptions imply that

$$E[y(i,k)y^*(m,l)] = K_G(i-m,k)\delta_{k,l}, \tag{5}$$

where $y(i,k) = y(i\Delta t - k\Delta\tau/2, \Delta\tau)$ and

$$K_G(n,k) = \sum_{m=-(I_{CR}-1)/2}^{(I_{CR}-1)/2} \sigma(m,k) e^{j2\pi mn\Delta f \Delta t}. \tag{6}$$

Let $\mathbf{y}_G(k)$ be the vector of samples from delay k

$$\mathbf{y}_G(k)^t = [\, y(0,k)\ \ y(1,k)\ \ y(2,k)\ \ \cdots\ \ y(G-1,k)\,]. \tag{7}$$

The covariance matrix for $\mathbf{y}_G(k)$ is a Toeplitz matrix $\mathbf{K}_G(k)$ whose i,m element is $K_G(i-m,k)$. The restriction imposed by the constraint of a target of finite extent is represented by Equation (6). That is, the only $\sigma(m,k)$ which can be nonzero are for $-(I_{CR}-1)/2 \le m \le (I_{CR}-1)/2$ and $0 \le k \le I_R-1$. This constraint on m is only meaningful if $\Delta f\, \Delta t I_{CR}$ is less than one and it limits the $\mathbf{K}_G(k)$ possible. If $\Delta f\, \Delta t = 1/N$, then the $\sigma(m,k)$ may be thought of as samples of the spectrum of the periodic extension of the covariance matrix at delay k. Only I_{CR} ($I_{CR} < N$) of the spectrum samples at each delay are nonzero.

We now introduce larger vectors and matrices so that the results which follow may be written compactly. The $\mathbf{y}_G(k)$ are loaded into one $GI_R \times 1$ vector $\mathbf{y}_G$. The covariance of $\mathbf{y}_G$ is a block diagonal

matrix $\mathbf{K}_G$ whose k^{th} diagonal block is $\mathbf{K}_G(k)$. Define $\Sigma(k)$ to be the $I_{CR} \times I_{CR}$ diagonal matrix

$$\Sigma(k) = diag\,[\,\sigma(0,k)\;\;\sigma(1,k)\;\;\sigma(2,k)\;\;\cdots\;\;\sigma(-2,k)\;\;\sigma(-1,k)\,] \tag{8}$$

and define Σ to be the $I_R I_{CR} \times I_R I_{CR}$ block diagonal matrix whose k^{th} block is $\Sigma(k)$. Then we have

$$\mathbf{K}_G(k) = \mathbf{W}_G^{\dagger} \mathbf{J}_{CR}^{t} \Sigma(k) \mathbf{J}_{CR} \mathbf{W}_G \tag{9}$$

where $\mathbf{W}$ is an $N \times N$ normalized DFT matrix ($N > G, N > I_{CR}$), $\mathbf{W}_G$ consists of the first G columns of $\mathbf{W}$, and

$$\mathbf{J}_{CR} = \begin{bmatrix} \mathbf{I}_1 & 0 & 0 \\ 0 & 0 & \mathbf{I}_2 \end{bmatrix} \tag{10}$$

where $\mathbf{J}_{CR}$ is $I_{CR} \times N$, $\mathbf{I}_1$ is an identity matrix of dimension $(I_{CR}+1)/2$, and $\mathbf{I}_2$ is an identity matrix of dimension $(I_{CR}-1)/2$. $\mathbf{J}_{CR}$ implements the assumption that some spectrum samples are zero. Having now defined several matrices, with a little more notation we can give a more precise relationship between the matrix Σ and the matrix $\mathbf{K}_G$. Let $\mathbf{M}_{CR}$ be the $I_R I_{CR} \times I_R N$ block diagonal matrix each of whose I_R blocks is $\mathbf{J}_{CR}$. Let $\mathbf{M}_R$ be the $I_R N \times I_R G$ block diagonal matrix each of whose I_R blocks is $\mathbf{J}_R$ ($\mathbf{J}_R$ is $N \times G$ and equals $[\,\mathbf{I}\;\;0\,]^t$). Finally, let $\hat{\mathbf{W}}$ be the $NI_R \times NI_R$ block diagonal matrix each of whose I_R blocks equals $\mathbf{W}$. Then

$$\mathbf{K}_G = \mathbf{M}_R^{t} \hat{\mathbf{W}}^{\dagger} \mathbf{M}_{CR}^{t} \Sigma \mathbf{M}_{CR} \hat{\mathbf{W}} \mathbf{M}_R . \tag{11}$$

Except for constraining some spectrum samples to be zero, if samples of $y_G(k)$ from each delay k were directly available, this problem would be very similar to that from [1] and the approach to the solution almost identical. Instead, for our problem, we have the data $\mathbf{r}$ which is described below.

Radar data. In a radar system, each signal is typically described as the product of a baseband signal times a complex exponential at the carrier frequency. All of the interactions of interest for narrow band radar systems may be described in terms of these complex valued baseband signals which will be called complex envelopes of the signals or simply the signals.

Let $s_T(t)$ and $s_R(t)$ be the complex envelopes of the transmitted and received signals, respectively. The distance of a point on the target from the transmitter/receiver is measured in time units as the two-way delay. The received signal is the superposition of the reflectivity from all two-way delays times the appropriate transmitted signals,

$$s_R(t) = \int_{-\infty}^{\infty} s_T(t-\tau) y(t-\tau/2,\tau) d\tau . \tag{12}$$

In Equation (12), the $\tau/2$ arises because it takes that long for the transmitted signal to get to a point at two-way delay τ. The available data are the sum of the radar return signal and additive white Gaussian noise

$$r(t) = s_R(t) + w(t)$$

$$r(t) = \int_{-\infty}^{\infty} s_T(t-\tau) y(t-\tau/2,\tau) d\tau + w(t) \;. \tag{13}$$

This equation forms the basis for our model. We now assume that samples of the data are available and we wish to estimate the covariance of the reflectivities from patches on the target. The sampled version of Equation (13) is

$$r(k) = \sum_{n=0}^{I_R-1} s_T(k\Delta t - n\Delta\tau) y(k\Delta t - n\Delta\tau/2, n\Delta\tau) + w(k) \;. \tag{14}$$

Given G samples $r(k)$, $0 \le k \le G-1$, we may write Equation (14) in vector form as

$$\mathbf{r} = \mathbf{S}_T^{\dagger} \mathbf{y} + \mathbf{w} \;, \tag{15}$$

where $\mathbf{r}$ is the vector of samples $r(k)$; $\mathbf{w}$ is a white noise vector with covariance $N_0\mathbf{I}$; $\mathbf{y}$ is the $GI_R \times 1$ vector of samples of the reflectivity of the target described above; and $\mathbf{S}_T^{\dagger}$ is a $G \times GI_R$ matrix composed of I_R

$G \times G$ submatrices each of which is diagonal:

$$\mathbf{S}_T^\dagger = [\, \mathbf{S}_0 \;\; \mathbf{S}_1 \;\; \mathbf{S}_2 \;\; \cdots \;\; \mathbf{S}_{I_R-1} \,], \tag{16}$$

where

$$\mathbf{S}_n = diag\,[\, s_T(-n\,\Delta\tau) \;\; s_T(\Delta t - n\,\Delta\tau) \;\; s_T(2\Delta t - n\,\Delta\tau) \;\; \cdots \;\; s_T((G-1)\Delta t - n\,\Delta\tau) \,].$$

Note that Equation (15) could also be written in terms of samples of $c(f,\tau)$ as

$$\mathbf{r} = \Gamma^t \mathbf{c} + \mathbf{w}, \tag{17}$$

where $\mathbf{c}$ is an $I_R I_{CR} \times 1$ vector of samples of the reflectivity of the target arranged as I_R subvectors each of length I_{CR} of the samples from each delay and Γ^t is a $G \times I_R I_{CR}$ matrix each element of which is a product of a sample of s_T times a complex exponential. More explicitly, Γ is given by

$$\Gamma = \mathbf{M}_{CR}\,\hat{\mathbf{W}}\mathbf{M}_R\,\mathbf{S}_T. \tag{18}$$

There are obviously issues associated with the appropriate or desired sampling rates of $r(t)$ and of $c(f,\tau)$. Some of these issues are addressed in [4]. The quality of the image obtained and the resolution achievable are intimately related to the sampling issues.

Since $\mathbf{r}$ is the result of linear operations on Gaussian random variables, $\mathbf{r}$ is a zero mean Gaussian random variable with covariance

$$\mathbf{K}_R = \mathbf{S}_T^\dagger \mathbf{K}_G \mathbf{S}_T + N_0\mathbf{I} = \Gamma^t \Sigma \Gamma + N_0\mathbf{I}, \tag{19}$$

Maximum likelihood solution. The loglikelihood function for the data is

$$L(\mathbf{K}_R;\mathbf{r}) = -\ln(\det\mathbf{K}_R) - \mathbf{r}^t(\mathbf{K}_R)^{-1}\mathbf{r}. \tag{20}$$

As shown first by Burg, et al. [5] a necessary condition for a matrix to maximize the loglikelihood function is the trace condition

$$tr\,[(\mathbf{K}_R^{-1}\mathbf{r}\mathbf{r}^t\mathbf{K}_R^{-1} - \mathbf{K}_R^{-1})\delta\mathbf{K}_R] = 0. \tag{21}$$

The matrix $\delta\mathbf{K}_R$ is a variational matrix which takes values in all possible additive variations of the matrix $\mathbf{K}_R$. The set of possible $\mathbf{K}_R$ is described by Equation (19). If we rewrite the trace condition in terms of $\mathbf{K}_G$, then we get

$$tr\,[\mathbf{S}_T(\mathbf{S}_T^\dagger\mathbf{K}_G\mathbf{S}_T + N_0\mathbf{I})^{-1}(\mathbf{r}\mathbf{r}^t - \mathbf{S}_T^\dagger\mathbf{K}_G\mathbf{S}_T - N_0\mathbf{I})(\mathbf{S}_T^\dagger\mathbf{K}_G\mathbf{S}_T + N_0\mathbf{I})^{-1}\mathbf{S}_T^\dagger\delta\mathbf{K}_G] = 0. \tag{22}$$

Equation (22) shows the three ways in which the present spectrum estimation problem differs from that in [1]. First, there is the signal matrix $\mathbf{S}_T$ appearing in (22) multiplying $\mathbf{K}_G$ wherever it appears. Second, the additive noise manifests its influence in the equation above through the appearance of $N_0\mathbf{I}$. Third, the matrix $\mathbf{K}_G$ in Equation (22) is a block diagonal matrix with Toeplitz blocks, not merely Toeplitz. Despite these differences an EM algorithmic approach to the solution can be derived in much the same way as in [1].

For each delay k, define the complete data vector $\mathbf{y}_N(k)$ to be the periodic extension of the data vector $\mathbf{y}_G(k)$

$$\mathbf{y}_N(k)^t = [\, \mathbf{y}_G(k)^t \;\; \mathbf{y}_A(k)^t \,] \tag{23}$$

where $\mathbf{y}_A(k)$ is an $(N-G)\times 1$ vector which augments $\mathbf{y}_G(k)$ to obtain a full period sample of the periodic process. Let $\mathbf{y}_N$ be the $NI_R \times 1$ vector made by stacking the $\mathbf{y}_N(k)$ to form one long vector. The complete data loglikelihood is

$$L(\mathbf{K}_N;\mathbf{y}_N) = -(\ln(\det\mathbf{K}_N)) - \mathbf{y}_N^\dagger(\mathbf{K}_N)^{-1}\mathbf{y}_N, \tag{24}$$

where $\mathbf{K}_N$ is a block diagonal matrix with each of the I_R blocks being a circulant Toeplitz matrix. Premultiplying $\mathbf{K}_N$ by $\hat{\mathbf{W}}$ and postmultiplying by $\hat{\mathbf{W}}^t$ yields a diagonal matrix. This diagonal matrix consists of the samples of the spectrum, some of which are constrained to be zero by the assumption of a

target of finite extent. This constraint results in

$$\mathbf{K}_N = \hat{\mathbf{W}}^\dagger \mathbf{M}_{CR}^t \Sigma \mathbf{M}_{CR} \hat{\mathbf{W}}. \tag{25}$$

What this implies is that after rotating the data $\mathbf{y}_N$ using the orthogonal matrix $\hat{\mathbf{W}}$, some of the entries of the resulting vector are zero. These elements may be removed from consideration by multiplying the resulting data by $\mathbf{M}_{CR}$, giving a vector of uncorrelated samples of a process whose covariance matrix is given by Σ

$$\mathbf{c} = \mathbf{M}_{CR} \hat{\mathbf{W}} \mathbf{y}_N \tag{26}$$

$$E[\mathbf{c}\mathbf{c}^\dagger] = \Sigma \ . \tag{27}$$

This vector $\mathbf{c}$ is the same as in Equation (17). Under the assumptions stated, the matrix $\mathbf{K}_N$ is not invertible. The reflectivity, however, almost surely does not have a component in the null space of $\mathbf{K}_N$ so we can make some sense of the complete data loglikelihood (24). A more correct way to write the complete data loglikelihood is in terms of the rotated coordinates $\mathbf{c}$ and its covariance Σ:

$$L(\Sigma;\mathbf{c}) = -\sum_{k=0}^{I_r-1} \sum_{i=-(I_\alpha-1)/2}^{(I_\alpha-1)/2} [\ln\sigma(i,k) + \frac{|c(i,k)|^2}{\sigma(i,k)}] \ . \tag{28}$$

The EM algorithm is an iterative algorithm which at each step updates the estimate for the Σ by maximizing the conditional expected value of the complete data loglikelihood over Σ. The E-step of the EM algorithm performs the expected value of (28) given the incomplete data and the previous estimate for Σ. The M-step consists of maximizing the result of this expectation over the $\sigma(i,k)$. Since the complete data loglikelihood in rotated coordinates separates into the sum of independent samples in (28), the result of taking the maximum over the spectral values at step p+1 is just

$$\sigma^{(p+1)}(i,k) = E[|c(i,k)|^2 | \Sigma^{(p)}, \mathbf{r}]. \tag{29}$$

The estimate of the covariance of $\mathbf{y}_N$ at step p+1 is found by transforming back to those coordinates

$$\mathbf{K}_N^{(p+1)}(l-n,k) = \frac{1}{N} \sum_{m=0}^{N-1} E[y_N(m,k) y_N^*(<m+l-n>_N,k) | \mathbf{r}, \mathbf{K}_N^{(p)}]. \tag{30}$$

This equation makes sense intuitively. It says that to find the maximum likelihood estimate over the constrained set of Toeplitz covariances, augment the covariance matrix at each delay with the conditional mean and mean square estimates of the missing lags. At the convergence point of the algorithm, the covariance estimates equal the conditional mean estimates of the lag products.

Returning to Equation (27), the estimate at step p+1 of Σ is given by the diagonal elements of the conditional expectation of $\mathbf{c}\mathbf{c}^\dagger$ or

$$\Sigma^{(p+1)} \stackrel{d}{=} E[\mathbf{c}\mathbf{c}^\dagger | \Sigma^{(p)}, \mathbf{r}], \tag{31}$$

where $\stackrel{d}{=}$ means that the off diagonal terms on the left are zero and the diagonal terms on the left equal the diagonal terms on the right. The computation indicated in (31) is a standard problem in estimation theory. This equation may be written as [4]

$$\Sigma^{(p+1)} \stackrel{d}{=} \Sigma^{(p)} + \Sigma^{(p)} \Gamma (\Gamma^\dagger \Sigma^{(p)} \Gamma + N_0 \mathbf{I})^{-1} (\mathbf{r}\mathbf{r}^\dagger - \Gamma^\dagger \Sigma^{(p)} \Gamma - N_0 \mathbf{I}) (\Gamma^\dagger \Sigma^{(p)} \Gamma + N_0 \mathbf{I})^{-1} \Gamma^\dagger \Sigma^{(p)}. \tag{32}$$

Equation (32) defines the iteration sequence used by the computations. The diagonal elements of Σ are the values which are displayed as the scattering function image of the target. Thus, at each stage an image is calculated and may be displayed. Some appropriate stopping criterion is used to terminate the algorithm. It should be pointed out that at each stage of the algorithm the conditional mean estimate of the reflectance is also generated. At step p, this estimate is

$$E[\mathbf{c} | \Sigma^{(p)}, \mathbf{r}] = \Sigma^{(p)} \Gamma (\Gamma^\dagger \Sigma^{(p)} \Gamma + N_0 \mathbf{I})^{-1} \mathbf{r}. \tag{33}$$

The magnitude or the magnitude squared of c are commonly viewed as images of the target. Thus both types of radar image commonly viewed are generated by our algorithm. We feel this is a unique feature of our algorithm.

Let the corresponding sequence of covariance matrices for the data $\mathbf{r}$ be denoted $\mathbf{K}_{\mathbf{r}}^{(p)} = \Gamma^{\dagger}\Sigma^{(p)}\Gamma + N_0\mathbf{I}$. Since this iteration is an EM algorithm, it has all of the properties associated with this type of algorithm. In particular, the incomplete data loglikelihood is nondecreasing in the sequence of covariance matrices $\mathbf{K}_{\mathbf{r}}^{(p)}$.

This section has presented a derivation of the equations used to produce target images. This approach starts with a model which accurately accounts for the random nature of radar reflections and adopts the maximum likelihood method of statistics to estimate delay-Doppler high resolution images of radar targets. Questions associated with uniqueness of spectrum estimates and convergence of the EM algorithm are discussed in the following section.

Convergence issues. This section addresses some of the convergence questions associated with the the EM algorithms proposed in earlier sections. These results are stated so that they apply to both the radar imaging problem studied here and the Toeplitz estimation problem from [1]. Let the integer M stand for $I_R I_{CR}$ in the radar problem. The matrix Γ refers to the Γ from the last section or simply to W_G if we are discussing the original Toeplitz problem. Also, N_0 is zero for the original Toeplitz problem.

One question of importance is the uniqueness of estimates of the parameters of interest. This question is addressed by looking at the Cramer-Rao bounds on the variance of estimates. The Cramer-Rao bounds are obtained by inverting the Fisher information matrix. When the Fisher information matrix is singular, these bounds are infinite. It is shown how singularity of the Fisher information matrix corresponds to nonuniqueness of parameter estimates.

Definition: Let γ_k denote the k^{th} row of the $M \times G$ matrix Γ. The $M \times G^2$ matrix F has k^{th} row given by $\gamma_k \otimes \gamma_k^*$, where $\otimes$ denotes the kronecker product.

Theorem 1: *The Fisher information matrix for estimating Σ given data r is equal to*

$$F\left(K_{\tilde{R}}^{-1} \otimes K_{\tilde{R}}^{-\dagger}\right)F^{\dagger} \ . \tag{34}$$

Proof: The Fisher information matrix is just the negative of the expected value of the second derivative of the log-likelihood function. This second derivative is evaluated in the appendix of [4] and taking the expected value yields the above expression.

Theorem 2: *Suppose the Fisher information matrix in (34) is singular and that the matrix K_R is positive definite. Then there does not exist a Σ which is positive definite which yields a unique maximum of the log-likelihood.*

Proof: Since the rank of the matrix $K_{\tilde{R}}^{-1} \otimes K_{\tilde{R}}^{-\dagger}$ equals G^2, and by the form of the matrix, the Fisher information matrix is singular if and only if the matrix F has rank less than M if and only if there exists a real vector s such that $F^{\dagger}s = 0$. Such an s exists if and only if there exists a real diagonal matrix D ($D = diag(s)$) such that

$$\Gamma^{\dagger}(\Sigma + \alpha D)\Gamma = \Gamma^{\dagger}\Sigma\Gamma \tag{35}$$

for all real α. If Σ is positive definite and maximizes the log-likelihood, then there exists a β such that for all $0 \le \alpha \le \beta$ the matrix $\Sigma + \alpha D$ is nonnegative definite and yields the same covariance matrix and hence the same value for the log-likelihood.

Corollary 1: For the spectrum estimation problem from [1], there does not exist a positive definite Σ which yields a unique maximum of the log-likelihood if $N > 2G-1$.

Proof: The matrix F constructed above has rank less than or equal to $2G-1$.

Note that this theorem does not say that the estimate of the Toeplitz covariance matrices generated by the algorithm are not unique. The theorem and its corollary relate to the uniqueness of the spectrum samples. For some problems the parameters of interest are in the covariance matrix $\mathbf{K}_R$ or in the Toeplitz matrix $\mathbf{K}_G$. There could be (and indeed are for M large enough) many Σ which yield the same estimate for $\mathbf{K}_R$. Theorem 2 can be applied to problems where one desires to know how big to make N. In general, it is desired to have N as short as possible in order to reduce the number of parameters to be estimated. If it can be shown that with a given N there exists a positive definite matrix which yields the maximum likelihood estimate for the covariance matrix and the conditions of the theorem are satisfied, then one

might consider using a smaller N to reduce the size of Σ.

For some problems, including the radar problem, it is the matrix Σ which is of interest. For these problems, it is very important to know when the estimate is unique. The radar imaging problem is one example. In the radar problem, the result which is displayed as an image is an array whose elements are the diagonal entries from Σ. In order to be able to generate a unique image, the conditions of theorem 2 must be satisfied.

Some of the issues associated with convergence of the EM algorithm for the problems described are addressed next. The following material is adapted from the material in [1] to be applicable to the radar problem as well.

Definition: Let $\mathbf{K}_R^{\beta}$ be the set of positive definite matrices $\mathbf{K}_R$ given by Equation (19) whose entries are bounded by some number a. Let $\overline{\mathbf{K}}_R^{\beta}$ be the closure of $\mathbf{K}_R^{\beta}$, the set of nonnegative definite matrices of this parameterized form.

The important issues for the following theorems are not the matrices which go into Γ. What is important is that any $\mathbf{K}_R \in \mathbf{K}_R^{\beta}$ may be written as

$$\mathbf{K}_R = \sum_{k=0}^{M-1} \sigma(k) \gamma_k^{\dagger} \gamma_k + N_0 \mathbf{I}, \tag{36}$$

where γ_k is the k^{th} row of Γ which is fixed once the model is specified. The only parameters which must be found are the $\sigma(k)$ which are specified to be greater than or equal to zero. An element of $\overline{\mathbf{K}}_R^{\beta}$ must have N_0 equal to zero. This is an important case for which we wish to guarantee that the estimate of $\mathbf{K}_R$ is nonsingular. Clearly for any fixed b the set $\overline{\mathbf{K}}_R^{\beta}$ is compact.

Theorem 3: *Let Γ be any fixed $M \times G$ matrix with complex entries. Let* $\mathbf{r}$ *be an observation of a $G \times 1$ 0-mean Gaussian random vector whose covariance is some positive definite hermitian symmetric matrix. Then*

a) There does not exist a singular $\mathbf{K}_R \in \overline{\mathbf{K}}_R^{\beta}$ such that $\mathbf{r}$ is in the range space of $\mathbf{K}_R$, with probability one.

b) The log-likelihood function is bounded from above over the set $\overline{\mathbf{K}}_R^{\beta}$, with probability one.

Proof: a) Suppose that any G rows of Γ are linearly independent. Since $\mathbf{K}_R$ is given by (36), a singular matrix in this class must be given by

$$\mathbf{K}_R^{\beta} = \sum_{k \in J} \sigma(k) \gamma_k^{\dagger} \gamma_k, \tag{37}$$

where $J \subset \{0,1,2,\ldots,M-1\}$ consists of $G-1$ or fewer integers which correspond to the nonzero diagonal entries of Σ. Since the true covariance for $\mathbf{r}$ is nonsingular, the probability that $\mathbf{r}$ lies in the subspace spanned by $\{\gamma_k^{\dagger} | k \in J\}$ is zero. Since there are a finite number of such spaces and the probability that $\mathbf{r}$ is in any one of them is zero, the probability that $\mathbf{r}$ is in the range of any singular $\mathbf{K}_R$ in the set is zero. If any G rows are not independent, then singular matrices may be written as the sum of more than $G-1$ outer products $\gamma_k^{\dagger} \gamma_k$. But the data would still have to lie in a subspace spanned by fewer than G independent vectors and thus the probability of this is zero and this part of the theorem follows.

b) This part follows from [6] where it is shown that the log-likelihood function is bounded above when the data are not in the range space of a singular covariance matrix in the set in question. The proof is based on the following facts. First, if $\mathbf{K}_R$ is nonsingular and its eigenvalues are bounded from above and below, the log-likelihood is bounded. Second, if $\mathbf{K}_R$ is singular and the data are in its range, the log-likelihood is unbounded above; but this is a zero probability event. Third, if $\mathbf{K}_R$ is singular (with rank n) and the data are not in its range, the log-likelihood is unbounded from below. This is shown by writing K_R as the limit as $\varepsilon \to 0$ of $\mathbf{K}_R + \Gamma^{\dagger}(\varepsilon \mathbf{I})\Gamma$. We examine the loglikelihood (20) when $\mathbf{r}$ is not in the range of $\mathbf{K}_R$. The term $-\ln(\det(\mathbf{K}_R + \varepsilon \Gamma^{\dagger}\Gamma))$ has bounded terms and an unbounded component of the form $-(G-n)\ln(\varepsilon)$. The term $-\mathbf{r}^{\dagger}(\mathbf{K}_R + \varepsilon \Gamma^{\dagger}\Gamma)^{-1}\mathbf{r}$ has bounded terms and an unbounded component of the form $-c/\varepsilon$, where c is some positive number. Then,

$$\lim_{\varepsilon \to 0} -(G-n)\ln(\varepsilon) - \frac{c}{\varepsilon} = -\infty \ . \tag{38}$$

There are more facts needed to prove convergence. One fact is that the log-likelihood is increasing at each step of the algorithm. Another is that the iterates stay in a bounded set so that the above theorems

apply. The theorems before this point apply to the problem no matter what algorithm is used to find the maximum likelihood estimate, while the theorems that follow apply for our particular algorithm.

Theorem 4: *The iterates defined by the EM algorithm (32) produce a sequence of log-likelihoods which are nondecreasing,*

$$L(\mathbf{K}_{\mathcal{N}}^{(p+1)};\mathbf{r}) - L(\mathbf{K}_{\mathcal{N}}^{(p)};\mathbf{r}) \geq Q(\mathbf{K}_{\mathcal{N}}^{(p+1)} \mid \mathbf{K}_{\mathcal{N}}^{(p)}) - Q(\mathbf{K}_{\mathcal{N}}^{(p)} \mid \mathbf{K}_{\mathcal{N}}^{(p)}) \geq 0 \ ,$$

where $L(.;.)$ is the log-likelihood for the problem.

Proof: This is just a result of the sequence being generated by an EM algorithm [7, 8].

Theorem 5: $L(\mathbf{K}_{\mathcal{N}}^{(p+1)};\mathbf{r}) = L(\mathbf{K}_{\mathcal{N}}^{(p)};\mathbf{r})$ *if and only if* $\Sigma^{(p+1)} = \Sigma^{(p)}$.

Proof: This is a result of the concavity of the complete data log-likelihood. Take the second derivative with respect to the variable being maximized, Σ. For each diagonal entry of Σ this derivative is either positive or zero. It's zero if and only if the previous corresponding entry of Σ is zero, and this entry would remain zero. Thus the maximizer is unique and is given by $\Sigma^{(p+1)}$. By the inequality from the theorem 3, theorem 4 follows.

The one last theorem we would like to have is that the iterates remain in a bounded set. It has been our experience in computations that the iterates do remain bounded, but we have had trouble proving this in the general case. We have observed in computations that Σ may tend to a singular limit. This is not precluded by any of the above theorems. In fact, for our radar imaging problem we do not wish to exclude this possibility since a zero estimate of the power reflected from a point simply means that there is no target at that point.

Conclusions. We have presented an algorithm for generating images of radar targets in the delay-Doppler plane. The approach has been estimation based because of our assumption of targets which are rough on the order of a wavelength of the carrier frequency. Some of the theoretical properties of this approach and uniqueness of estimates have been discussed. Presently we are implementing the proposed algorithm and performing computational studies. We are also addressing several theoretical issues. One issue of particular importance is the incorporation of specular components in the algorithm. These points would have a different statistical characterization than the diffuse components considered here and a correspondingly altered loglikelihood to be maximized. Computationally, we are examining the convergence of our algorithm, its computational complexity, and comparing the performance of the algorithm to other approaches.

REFERENCES

References

1. M. I. Miller and D. L. Snyder, "The Role of Likelihood and Entropy in Incomplete-Data Problems: Applications to Estimating Point-Process Intensities and Toeplitz-Constrained Covariances," *Proceedings of the IEEE*, vol. 75, pp. 892-907, July 1987.
2. H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part III*, Wiley and Sons, New York, NY, 1971.
3. J. Shapiro, B. A. Capron, and R. C. Harney, "Imaging and Target Detection with a Heterodyne-Reception Optical Radar," *Applied Optics*, vol. 20, pp. 3292-3313, October 1981.
4. D. L. Snyder, J. A. O'Sullivan, and M. I. Miller, "The Use of Maximum-Likelihood Estimation for Forming Images of Diffuse Radar-Targets from Delay-Doppler Data," *IEEE Transactions on Information Theory*, vol. to appear 1989.
5. J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of Structured Covariance Matrices," *Proceedings of the IEEE*, vol. 70, pp. 963-974, September 1982.
6. D. R. Fuhrmann and M. I. Miller, "On the Existence of Positive Definite Maximum-Likelihood Estimates of Structured Covariance Matrices," *IEEE Transactions on Information Theory*, vol. 34,

July 1988.

7. A. D. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Society*, vol. B39, pp. 1-37, 1977.

8. C. F. J. Wu, "On the Convergence Properties of the EM Algorithm," *The Annals of Statistics*, vol. 11, pp. 95-103, 1983.